DOCTOR OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES

Epitranscriptomic analyses of the dihydrouridine RNA modification in Bacteria and Eukarya

Finet, Olivier

Award date:
2018

Awarding institution:
University of Namur

Epitranscriptomic analyses of the dihydrouridine RNA modification in Bacteria and Eukarya

Olivier Finet
FRIA fellowship
Laboratory of Molecular Genetics, URPhyM
Faculty of Medicine, University of Namur

Public thesis defense
7th December 2018

Members of the jury :
- Pr. O. De Backer (President, UNamur)
- Pr. N. Gillet (UNamur)
- Dr. M. Graille (Ecole polytechnique, Université Paris-Saclay)
- Dr. R. Hallez (UNamur)
- Dr. D. Hermand (Secretary, UNamur)
- Pr. D. Lafontaine (ULB)

Université de Namur
Faculté de Médecine

I'll be fine - if you give me a minute, a man's got a limit. I can't get a life if my heart's not in it.

Oasis

Science ultimately is the act of discovering new knowledge and when you learn about science in school, you're not actually doing science, you're learning the knowledge that people discovered previously and it's more like studying "history of science" […]. To actually get in a lab and to get your hands on experiments that are actually discovering new knowledge is so hard and it takes so long.

Craig Mello, 2006 Nobel Laureate

La science a une grosse probabilité de merder.

Xavier De Bolle

True perfection has to be imperfect. I know that that sounds foolish but it's true

Oasis

TABLE OF CONTENTS

ABSTRACT

A large catalogue of modified nucleosides is found within RNAs. Most of these chemical modifications have been known for the past fifty years. However, the first epitranscriptomic studies are recent and shed light to a new layer of gene expression regulation. To date, less than ten RNA modifications have been studied at the transcriptomic level due to the difficulty to determine their distribution at the single-nucleotide resolution.

The dihydrouridine RNA modification (D) is a product of the reduction of uridine by a conserved family of dihydrouridine synthases (Dus) and has been exclusively studied within the context of tRNA and rRNA biology. Despite its universal conservation among the three domains of life, little is known about the biological relevance of this modification.

Taking advantage of the previously reported fluorescent labeling of tRNAs at dihydrouridylated positions, we have developed D-seq (dihydrouridine-sequencing), a sensitive method to detect the transcriptome-wide distribution of dihydrouridines, which we confirmed with dot blot and primer extension assays. Remarkably, the D-seq pipeline could be easily applied to other RNA modifications that are sensitive to the rhodamine labeling and for which the cognate RNA modifying enzyme mutants are available.

Among the 372 putative dihydrouridylated sites found on the yeast transcriptome, 60% were located on well-known positions of tRNAs. Surprisingly, the remaining predicted sites were found onto messenger RNAs. The dissection of D distribution revealed an enrichment on coding sequences and a non-random distribution on codons. Unlike coding RNAs, hypodihydrouridylated tRNAs were remarkably more abundant than their modified counterparts. The conservation of this feature in another eukaryote is currently under investigation on a human transcriptome.

Interestingly, the *E. coli* coding transcriptome was devoid of any detectable dihydrouridine in the tested condition, implying that mRNA dihydrouridylation could be eukaryotic-specific. Our data also suggest that a novel yet uncharacterized bacterial Dus enzyme modifies the 23S rRNA and might target other RNA species.

Altogether, our data demonstrate the reliability of the unbiased D-seq technique to map the dihydrouridine at the single-nucleotide resolution. Moreover, we propose dihydrouridine as a potential new internal mRNA modification in Eukarya.

There's an urgent and high demand for additional technology developments for all kinds of RNA modifications.

Tao Pan in (Chi, 2017)

ACKNOWLEDGMENTS

I would like to thank the members of the jury – Pr. Olivier De Backer (president), Pr. Nicolas Gillet, Dr. Marc Graille, Dr. Régis Hallez, Dr. Damien Hermand (secretary) and Pr. Denis Lafontaine – for the time they took not only to read this manuscript, but also to provide fruitful comments in order to improve the thoroughness of my work.

It is essential to thank all the people who participated in the experimental setup, in data generation and analysis. This manuscript would have never existed without their involvement: Dr. Damien Hermand – supervisor of this work – who led the project with exciting ideas and unwavering commitment; Carlo Yague-Sanz who performed NGS analyses with an outstanding motivation and professionalism; Dr. Antonin Morillon and Dr. Maxime Wéry who shared their expertise on large-scale techniques to make all NGS possible; Pr. Denis Lafontaine who kindly accepted to visit us – along with Pr. Henri Grosjean – in order to discuss about my project and who subsequently welcomed me in his laboratory; Dr. Félix Ernst who generated human cell lines and HPLC profiles; Valérie Migeot for GST-purified proteins; Sarah Legraie who worked on the *dus4* project that is only partially evoked here; Dr. Phong Tran and Lara Kruger who looked at the yeast cytoskeleton in our mutants of interest.

During the course of this thesis, my work was also evaluated by a committee composed of Pr. Benoît Muylkens and Dr. Lidia Vasilieva; thanks for their time and interest.

Figures 21, 23, S15 and S17 were provided by Lafontaine lab (ULB, BE); Fig 25 was provided by Tran lab (Institut Curie, FR); Fig. 28 was generated thanks to fruitful discussions with R. Hallez (UNamur, BE).

Last but not least, I am indebted to the past and present members of the GéMo team: Damien and Carlo who are already mentioned above; Julie, Valérie and Philippe who taught me how to work with *S. pombe*; Clément and Fanélie who always helped me to solve all technical and protocol issues; all the bachelor and master students who broadened our happy team.

Have a good read!

REMERCIEMENTS

Une thèse, ce ne sont au final que quatre années d'une vie. Mais n'est-ce pas mieux si celles-ci resteront un fabuleux souvenir? On ne parle pas de recherche ici, on parle de l'humain. La recherche, c'est un concentré de frustrations qui, de manière un peu sado, sont parfois excitantes. C'est ce qui nous fait tenir, cet espoir d'avoir un résultat (une fois tous les six mois, ça nous suffit au final… quoique !). Et puis, ce qui nous fait tenir – en tout cas moi – ce sont les gens qui nous entourent. Gandhi disait « un sourire ne coûte rien et produit beaucoup ». Je pense que c'est la même chose avec un merci. Alors, merci à vous ;

Tout d'abord Damien, bien évidemment. Mon « promoteur » comme on dit en Belgique, « chef » comme on l'appelle au labo. Mais pas un chef autoritaire, non, plutôt un leader à l'oreille attentive. De jeune biologiste fraîchement diplômé – sachant à peine ce qu'était une modification de l'ARN – tu as fait de moi un scientifique qui a confiance en sa recherche et sa connaissance, tout en ne nous faisant jamais oublié que nous n'étions au final tous que bien ignorants par rapport à la connaissance du monde. Et savoir qu'on ne sait pas, c'est une philosophie qui me servira toute ma vie. Je n'ai pas dû être un thésard facile tous les jours – on se souviendra de cet interview à la RTBF où j'ai légèrement fait du forcing ou de ma dépression presque chronique devant l'absence de phénotype de mon mutant –, mais tu as toujours su trouver le positif et je t'en remercie.

J'ai longtemps hésité à me réorienter dans la recherche en sciences sociales, pour étudier mes phénomènes de collègues. Mais au final, les litres de larmes de rire versées valaient mieux que n'importe quelle étude sociologique. Je me dois de commencer par Mamy Valoch – la plus scintillante et pétillante étoile de la galaxie comme elle a « exigé » que je l'appelle dans ces remerciements – sinon elle va déprimer (« pourquoi personne ne me comprend ? ») et noiera son chagrin dans des Krisprolls pâté crème (« je ne grignote jamais entre les repas »). Well, par quoi commencer ? Merci à toi, ô Premier Agent Spécialisé Principal Supérieur Avancé et tous les adjectifs du dictionnaire, pour ces citations bouddhistes, « t'y es déjà allé, toi, à l'Hôtel Sainte-Maxime ? », ces jeux du rond à ne plus finir, « c'était votre sœur », tes restes de pâtes bolos, gras de poulet et j'en passe, « qui êtes-vous ? », ces remix de Beethoven, « Sodome et Gomorrhe », etc. J'hésite même à te remercier pour tous ces U&P que tu n'as pas été chercher, pour m'avoir mis à bout exprès pour que je dise « Valooooooch », « Vieille charogne » ou « Valérie Migeot », pour m'avoir appris la meilleure recette de pesto-crème du monde, pour m'avoir montré avec tant d'intérêt la courbe d'évolution du prix au mazout, pour ne jamais retenir « Ste11, stE11, *sTE11*, oh broque de viol, comment on imprime Popi ? ». Oui, merci pour tout ça. Reste toujours aussi jeune dans ta tête, c'est ça qui compte et c'est ça qu'on aime ! Pour les suivants, je ferai pas ordre alphabétique, pas de jaloux ! Carlito, pardi ! Non seulement tu sais que j'ai toujours été impressionné (jaloux ?) de tes connaissances à la fois wet et dry – et d'ailleurs un merci infini pour ton apport essentiel (le mot est faible) et de qualité au projet dihydrouridine –, mais surtout pour ta Vishnu-attitude. J'espère que tu te rappelleras avec émotion de mes crises de nerf sur Excel et des légers tapotements sur le bureau ☺. Bonne humeur, sourire, volonté, amateur de bière, disponible, intéressé, voici le portrait non-exhaustif de notre Carli-Carlo – parfois

Carlocifer, mais ça nous faisait tellement rire – à qui je souhaite tout le meilleur au Canada… mais avec une personnalité aussi lumineuse que la tienne, je ne me fais aucun soucis ! Viens ensuite l'ami Consti, dont son amour pour les phrases répétitives (« j'ai faim », « pas plus d'douze nœuds hein », « coucou petite perruche ») est à la hauteur de son amour pour l'Orval. Notre passion commune pour le gras et Nicole Croisille, pour les cartons et le rikiki, ta passion pour mes oreilles et ma passion pour tes floches de pull ont fait une belle harmonie qui m'a bien manqué quand tu es devenu Dr. *elegans*. Heureusement, l'appel de l'alcool aidant, ça a continué en dehors du labo et je te remercie pour tous ces beaux souvenirs… D'ailleurs do you know El Hostel ? Ratusz ! Fanélie – « enchanté rôti moutarde » – ou notre fan inconditionnelle de l'humour bien sale ; non seulement c'était très enrichissant de partager une partie de ma thèse avec une postdoc qui a pu me prodiguer tous ses bons conseils, mais c'est surtout tout le reste qui me fait encore rire maintenant ; courir après un caddy sur le parking du Colruyt (love Collect&Go), visiter l'écluse de Jambes, faire des photos aussi débiles que nous pour distraire Miche-Miche, craquer pour l'alcool et le gras (love Quick et cannette sur le Grognon). Merci pour ta folle folie et pour ton ouverture d'esprit, ô petit bonhomme en mousse qui « turn around » ! Philippe-Paul, toujours le sourire et le positivisme… nous n'aurons été collègues que bien trop peu de temps et c'est dommage ! J'espère avoir fait honneur à ta paillasse en la reprenant et j'espère surtout qu'on va boire plein de bières à mon souper de thèse pour honorer la levure ; je compte sur toi :D!

Je n'oublie pas tous ces étudiants passés par GéMo, à savoir, par chronologie : Thomas, un mec intéressé et intéressant, facile à vivre et qui a fourni un effort sans limite pour comprendre notre humour ; Jérémy, certes complètement fou, mais qui nous aura bien fait rire avec son « 18 novembre » et ses rêves plus que surréalistes ; Sarah, ma seule mémo, que j'espère ne pas avoir trop traumatisée… malgré nos quelques divergences scientifiques, je suis ravi d'avoir passé ces quelques mois à t'apprendre les bases de la science et te remercie pour ton inflexible sympathie ; FX ou « Le Bio de l'ImproNam » qui m'a rappelé de beaux souvenirs et qui, surtout, aura supporté ma version du monde du métal ; Ysaline (Lebruuuuuun, à chanter sur l'air de « Sur un air latino ») qui, nous seulement, met ses tips un par un, mais surtout, a appris à ne plus me juger même si les Mouscronnois sont les carolos de Charleroi ; JB, le gars qui s'est fondu en GéMo en 2 minutes tellement il est taré, merci à lui d'avoir joué le jeu du « filet de poulet à la mayonnaise ».

Pour ceux qui n'auraient pas encore compris, je me lâche complètement et je paraphrase Carlo qui m'a dit « tes remerciements vont être plus longs que ta conclusion ». Bon ben, de toute façon, j'ai conclu en une page, donc je suis déjà foutu. Alors, continuons ! Nos voisins de l'URBM : merci Marbite pour ton amitié sans limite, pour aimer les petites choses de la vie sans aucune prétention, pour ton accueil toujours si jovial que ce soit dans ton bureau, à Oxford (bientôt mon tour !) ou à Loyers (mimimimimi) ; merci Katymouille de ne pas m'avoir détesté pour tout ce que je t'ai fait subir ; ce n'était qu'affection pour ta personnalité en or et j'ai hâte qu'on cause encore ensemble des heures pour ragoter (gloire à la belette) ; merci Aurore pour ton sourire rayonnant et communicatif, du Cercle aux soirées sans fin en salle radiologie, en passant par le match le plus épique de Belgique-Japon ; merci Patoche pour avoir été la rayon de soleil que j'attendais tous les matins, pour ton humour et pour tous les ragots des facs ; merci aux anciens comme Madame Iche, Eme, Doudou, Charlotte, Nayla ou Arnaud dont les innombrables fou rires partagés sont les meilleurs témoins

de mon affection pour vous ; merci Pierre pour ton optimisme et ton soutien moral indéfectibles quand la science fait chier ; merci Pauline pour ton franc-parler et pour m'avoir fait découvrir l'œuf parfait ; merci Kévin pour avoir toujours été attentif aux autres et pour la cool-attitude qui est la tienne ; thanks Vicky for your constant energy and your epic birthday parties (wish you the best in Sw…Norway); merci Bruce, même si t'es qu'un sale chômeur, pour vivre ta vie sans te prendre la tête et pour partager ta bonhomie sans retenue ; merci Agnès – the queen of the blind test – pour m'avoir toujours toléré avec sympathie dans ton bureau, quand j'étais en quête de bavardages ; merci Maffiou pour m'avoir diverti pendant ces horribles et interminables heures en pièce radioc et pour notre caoutchouc-adventure ; merci Jérôme (feu M. L'Assistant Coppine) pour avoir manié avec tant de bonheur la boîte à meuh ; merci à Gwen pour ton indéniable motivation à toujours partager les événements sociaux du labo ; merci Elie pour ta nonchalance pleine de jovialité ; merci Angy pour avoir si souvent ri de bon cœur à mes blagues ; merci à Xa d'être un PI si inspiré et inspirant.

Et puis il y a les plus fidèles des plus fidèles, ceux qui n'ont jamais jugé mes choix et mes plaintes, même s'ils ne comprenaient pas toujours ce que je faisais, m-e-r-c-i à vous : les vieux de Colisani (mes copains de la sieste et de la raclette, des jeux pas trop compliqués et du farniente apaisant depuis plus de 10 ans) ; Shmu et Biloute pour votre pouvoir presque magique à diffuser de la joie là où vous passez (on s'le fera ce méga voyage avec Laurent, on y arrivera :D) ; mes petiots Skål (franchement, c'est au-delà de l'amitié ce que je ressens pour vous) ; la Dream Team du sushi et de la bière (là aussi ça commence à faire un paquet d'années ; ma confiance en notre amitié est infinie et je sais que c'est vous que je dois appeler si je veux me marrer), les vieux survivants du Cercle Bio (pas moyen de s'ennuyer avec vous, c'est comme si nos défauts s'additionnaient pour s'annuler et faire un cocktail d'une perfection inimaginable). Merci, merci, merci !

Enfin, je le disais déjà dans mes remerciements de mémoire mais c'est la vérité ; qu'est-on sans une famille ? Merci à mes parents pour être le socle le plus solide sur lequel je puisse compter, pour avoir été si compréhensifs sur le fait que la recherche est plus souvent faite de bas que de hauts, pour me pousser à réaliser mes rêves ; merci à mes deux grandes sœurs qui, pour l'une a apporté avec mon cher beau-frère les deux plus merveilleuses sources de bonheur de notre famille – Will et Chacha – et pour l'autre a toujours porté attention à mes joies et frustrations avec beaucoup d'affection; merci à Grand-Maman d'être toujours le regard porté vers l'optimisme et pour toutes les belles leçons de vie que tu m'as transmises ; merci à mon Parrain et à Charlotte pour ces moments, le plus souvent devant un verre à notre QG Place Saint-Aubain, qui m'ont permis de m'échapper des tracas (souvent bien futiles) de chercheur.

Merci. Merci. Merci encore. Je vous dois tout.

INTRODUCTION

I. THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

A. Historical considerations

My own thinking (and that of many of my colleagues) is based on two general principles, which I shall call the Sequence Hypothesis and the Central Dogma.

(Crick, 1958)

Sixty years after the Francis Crick's statements on protein synthesis, the *Central Dogma* is still considered as the simplest and most accurate model to define the molecular biology. However, the Central Dogma alone cannot explain the flow of genetic information in living organisms. Contrary to what is commonly held in textbooks, the dogma does not mention that the nucleic acid sequence acts as a code for the amino acid sequence – this latter is called the *Sequence Hypothesis*. The Central Dogma should instead be defined as a negative statement implying that transfer of information originating from protein does not occur. The idea of a precise alphabet that permits the sequential flow of information from nucleic acids to proteins can be drawn by the combination of both principles (Crick, 1970). In this manuscript, and for the sake of convenience, the expression of *Central Dogma* will be used to refer to the *DNA → RNA → protein* transfer of information.

For decades, molecular biologists have studied the biochemical features of DNA, RNA and proteins and the biological processes that correlate them with one another (transcription and translation). Many breakthroughs considerably complexified the basic Central Dogma, leading to the investigation of unexpected layers of regulation (*e.g.* DNA methylation, RNA-dependent transcriptional interference, co-translational control of protein folding, *etc.*). Because the Central Dogma has been validated in all Domains of life, the interest of the scientific community for DNA-, RNA- and protein-related researches has never stopped growing.

B. RNA: a key player in the Central Dogma… and beyond

Finally, DNA appeared […]. RNA is then relegated to the intermediate role that it has today – no longer the centre of the stage, displaced by DNA and the more effective protein enzymes.

(Gilbert, 1986)

i. Generalities

The Central Dogma usually implies that the fate of a biological system is to perform a specific function (protein) from a well-organized source of information (DNA).

Although it is a true statement (with exceptions), it relegates RNA to a second role in which it is an intermediate molecule that connects information and function. However, RNA plays a pivotal role in the Central Dogma by potentially being at the origin of (molecular) life (point ii.) and by being tightly regulated (point iii.). Moreover, RNA can be far more than a coding messenger and displays numerous functions beyond the Central Dogma (point iv.).

ii. RNA in the story of life

When the first experimental evidence of a self-splicing RNA was published in 1982 – along with the revolutionary concept of *ribozyme* –, the idea that catalytic RNAs could be at the origin of life emerged (Kruger et al., 1982). One year later, Altman and colleagues described the conditions in which the RNase P enzymatic RNA cleaves a tRNA precursor (Guerrier-Takada et al., 1983). These discoveries, awarded by the 1989 Nobel Prize in Chemistry, led to the establishment of the *RNA world* hypothesis (Gilbert, 1986). In his one-page article, Gilbert exposed that if a unique molecule – such as RNA – can combine both the informational and catalytic properties, then an RNA world could have emerged from the prebiotic soup. He explained a five-step process in which; (i) RNA molecules replicate themselves, (ii) some daughter molecules acquire a function, (iii) RNA molecules undergo recombinations and mutations, which therefore widen the range of catalytic activities, (iv) with the development of RNA adapters that carry amino acids and a sort of ancestral ribosomal activity, the first proteins are synthesized, and (v) by reverse transcription, the more stable DNA replaces the RNA as a holder of the information. Altogether, these works pointed out RNA as the potential main character in the story of life.

iii. RNA-related regulations

At the biochemical level, RNA is a versatile polymer with molecular features ranging from its ability to form base pairs with other nucleic acids to its folding capacities (reviewed in (Geisler and Coller, 2013)). These characteristics make the RNA a highly dynamic and regulated biomolecule.

Intricate mechanisms act together to fine-tune the production of RNAs (transcriptional activity or synthesis through an RNA-dependent RNA polymerase) that are then monitored by RNA surveillance pathways. When RNA does not meet the quality standards, it is degraded by specific (*e.g.* nonsense-mediated decay) or more general (*e.g.* eukaryotic and archaeal exosomes or bacterial degradosome) machineries. The RNA production/degradation dichotomy ensures proper biological functions and will not be described here. Nevertheless, it is important to mention that RNA – this *simple intermediate* – is obviously at the core of numerous regulatory pathways.

a. An example of how RNA is regulated: tRNA QC

Due to the vital role of tRNAs in translation, cellular pathways exist to control their functional integrity. A quality control (QC) mechanism has been shown to act on an hypomethylated[1] form of the precursor $tRNA_i^{Met}$ (pre-$tRNA_i^{Met}$). The hypomethylation most likely leads to the misfolding of the transfer RNA (tRNA) that is

[1] The concept of modified RNA (*e.g.* methylation) is introduced thereafter (Introduction II.)

therefore polyadenylated by the Trf4 poly(A) polymerase, subunit of the TRAMP complex (Trf4/Air2/Mtr4 Polyadenylation complex). The Rrp6 3'-5' exoribonuclease is then recruited and degrades the hypomethylated pre-$tRNA_i^{Met}$ (Kadaba et al., 2004; Kadaba et al., 2006; Megel et al., 2015). This TRAMP-dependent decay that acts on a pre-tRNA is called the *nuclear surveillance pathway* (**Fig. 1A**).

Mature tRNAs also undergo degradation when they are hypomodified. Yeast mutants for nonessential RNA modifications displayed *rapid tRNA decay* (RTD) at 37°C (**Fig. 1B**) (Alexandrov et al., 2006). The degradation was then shown to depend on 5'-3' exoribonucleases such as Rat1 and Xrn1 (Chernyakov et al., 2008). Investigations on the molecular features of tRNAs targeted by RTD revealed that they are subjected to non-canonical 3'-CCACCA addition or 3'-polyadenylation (Wilusz et al., 2011).

iv. Beyond the Central Dogma: the noncoding world

Although they are noncoding, tRNAs and ribosomal RNAs (rRNAs) are associated with the Central Dogma because their most obvious function is to help decoding messenger RNAs (mRNAs). However, high-throughput transcriptional analyses revealed the existence of many more noncoding RNA species, with up to 85% of the human genome being transcribed – whereas the exome represents less than 2% of the genome (Djebali et al., 2012). This pervasive transcription generates short and long ncRNAs that are associated with specific genomic regions (telomeric transcripts, promoter-associated RNAs, *etc.*), that have well-established functions (small-interfering RNAs, long (intergenic) noncoding RNAs, *etc.*) and/or that are expressed in a specific physiological context (yeast meiotic transcripts, mammalian DNA damage RNAs, *etc.*) (Tisseur et al., 2011).

Cech and Steitz amazingly reviewed what they called *the noncoding RNA revolution* (Cech and Steitz, 2014). They emphasized that noncoding RNAs (ncRNAs) are not only involved in translation (as for rRNAs and tRNAs), but also in a wide variety of biological processes. The description of how noncoding RNAs can be functional is beyond the scope of this manuscript. However, two interesting examples are given;

a. MALAT1, a functional long noncoding RNA

Among the 16,000 long noncoding RNAs (lncRNAs) encoded by the human genome, MALAT1 (Metastatis-Associated Lung Adenocarcinoma Transcript 1) is one of the most abundant with about 3,000 copies/cell. MALAT1 is a 7kb-long RNA that is known to be processed by the RNase P ribonuclease, resulting in the formation of two fragments; (i) a short RNA that is exported to the cytoplasm and (ii) a long RNA that is nuclear-retained and localizes to the nuclear speckles (Wilusz et al., 2008) (**Fig. 2A**). Depletion of MALAT1 results in the alteration of SR splicing factors phosphorylation and in the variation of alternative splicing profiles of a subset of pre-mRNAs. The noncoding MALAT1 is therefore considered as a pre-mRNA processing element, albeit still elusive on its mode of action (Sun et al., 2017).

b. When cleaved tRNAs impact the ribosome biogenesis

tRNA-derived small RNAs (tsRNAs) are short ncRNAs formed by the cleavage of mature tRNAs, but are not nonsense by-products (Li et al., 2018) (**Fig. 2B**). In the

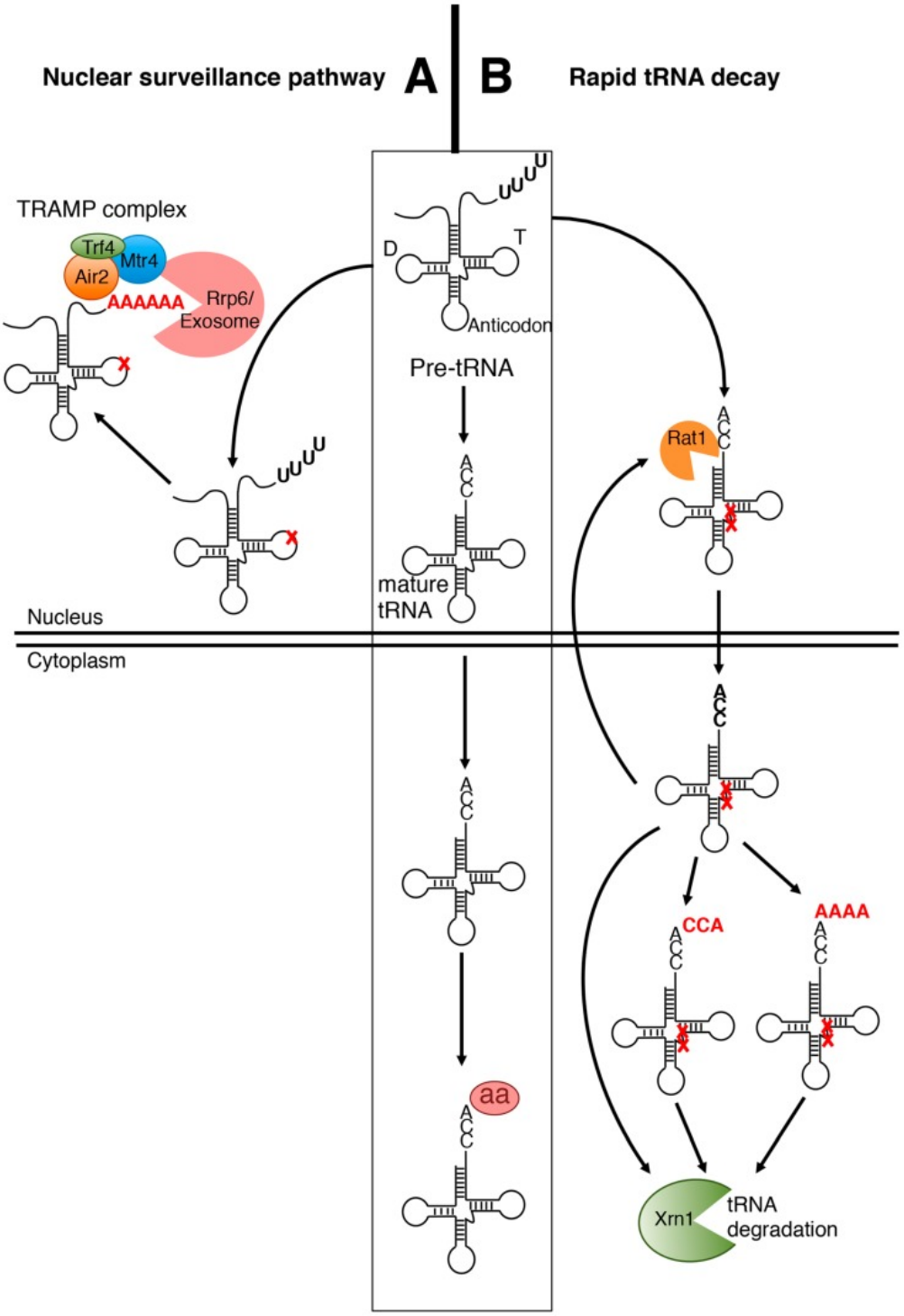


Figure 1 – tRNA quality control pathways. The normal tRNA maturation (framed in the centre) includes 5'- and 3'-trailers removal, 3'-CCA end addition, nucleocytoplasmic export and amino acid (aa in red) charging. **(A)** Nuclear surveillance pathway involves the pre-tRNA TRAMP-dependent polyadenylation and subsequent degradation by the nuclear exosome. **(B)** Rapid tRNA decay involves the nuclear or cytoplasmic degradation of (3'-modified) mature tRNAs. Red crosses highlight the lack of tRNA posttranscriptional modifications (Megel et al., 2015).

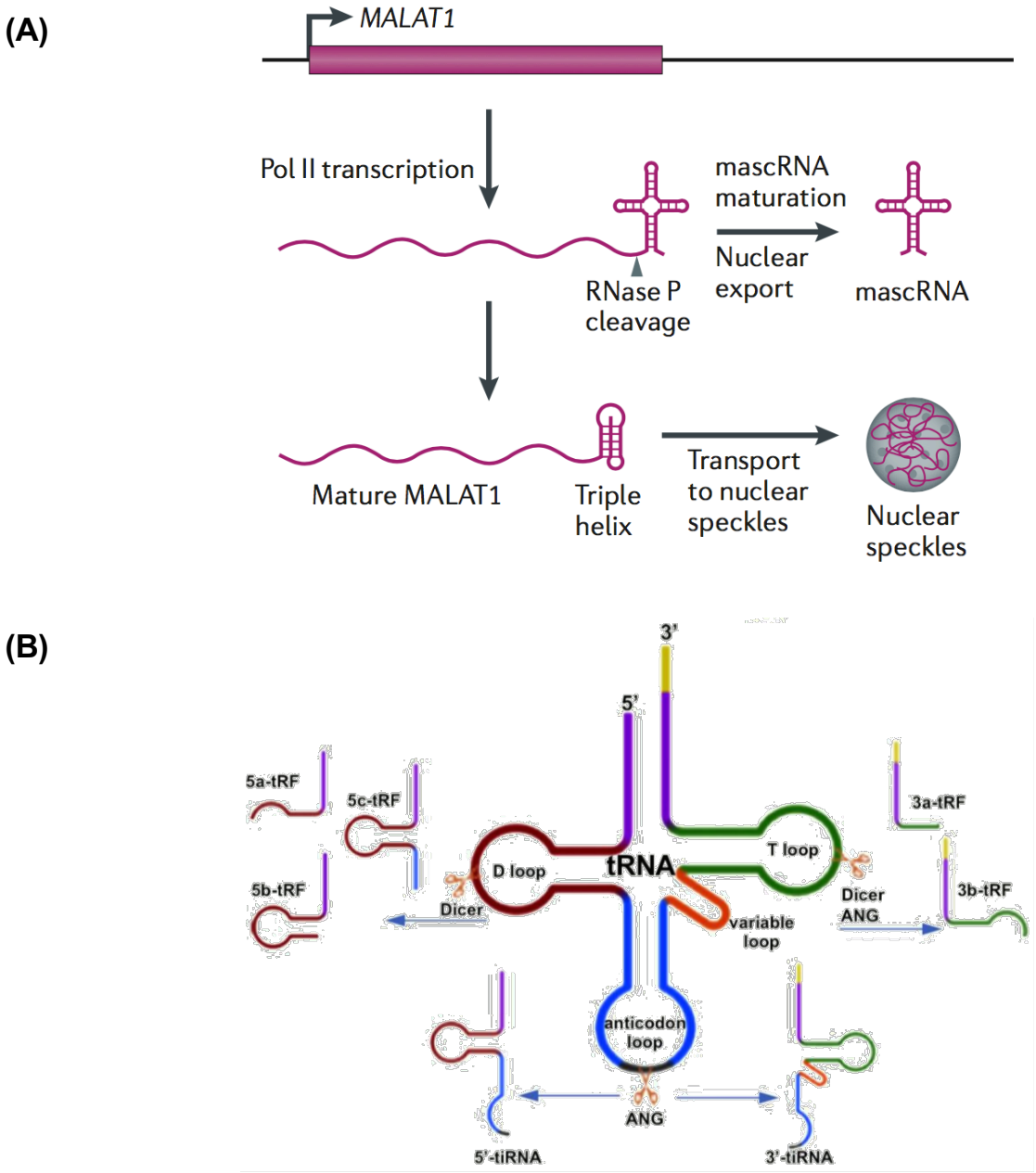


Figure 2 – Functional noncoding RNAs. (A) MALAT1 is a processed long noncoding RNA; a 61nt-long RNA (mascRNA) is exported to the nucleus and a 7kb-long RNA (mature MALAT1) adopts a stable 3'-triple helix to be localized to the nuclear speckles (Quinn and Chang, 2016). **(B)** tRNA-dervied small tRNAs are classified into two classes; tRf (tRNA-derived fragment) and tiRNA (tRNA-derived stress induced RNA). The cleavage is performed on different tRNA positions by ribonucleases such as Dicer or angiogenin (ANG) (Li et al., 2018).

ciliate *Tetrahymena thermophila*, 3'-tsRNAs have been shown to modulate the TXT complex formation (Ago/Piwi Twi12 protein; Xrn2 exonuclease; Twi-associated novel 1 Tan1 protein) by specifically interacting with one of its subunits (Twi12). Binding of the ncRNA therefore promotes the TXT nuclear localization where it acts as a pre-rRNA processing factor. Preventing TXT nuclear import induces ribosomal RNA biogenesis defect (Couvillion et al., 2012). This example cleverly shows how a functional ncRNA (tRNA) can be processed into another ncRNA (tsRNA), which in turn works as a regulator of RNA metabolism.

v. Conclusion

RNAs are functional; on the one hand they carry the information to the translational machinery to allow the protein synthesis (mRNAs) and on the other hand they tightly regulate a wide range of biological mechanisms (ncRNAs). Moreover, RNAs are subjected to many forms of regulation and some are even catalytically active (ribozymes). They can be both information and function, which makes them unique in the Central Dogma. In conclusion, it would be a terrible mistake to consider RNAs as simple molecular intermediates, trapped between the information and the function. RNA-related investigations are of high interest to help deciphering the mysteries of biology.

II. RNA MODIFICATIONS

All abbreviations used to name RNA modifications are found in the Modomics or the RNA modification databases (http://modomics.genesilico.pl/, http://mods.rna.albany.edu/) (Boccaletto et al., 2018; Cantara et al., 2011) and in the Annexes of this manuscript. For example, *U2-m^6Am_{30}* stands for the m^6Am modification occurring at position 30 of the U2 snRNA. In the nomenclature posttranscriptional modifications (PTMs), the letters written before the nucleotide refer to a modified nucleobase, whereas the annotation written after the nucleotide refers to a ribose modification. The m^6A**m** modification carries two methylations; on position 6 of the nucleobase and **on the ribose**. The numbering system that is applied to the nucleobase(s) and the ribose is depicted in **Fig. S1**.

A. Biochemistry of RNA modifications, diversity and biosynthesis

Now we know that 163 post-transcriptional modifications of RNA introduce a functional diversity that allows the four basic ribonucleotide residues to gain diverse functions.

(Boccaletto et al., 2018)

The RNA backbone is a succession of covalently bound ribonucleotides whose nucleobase, ribose or 5'-extremity can be modified. To date, there are more than 150 known RNA chemical modifications, spanning the three domains of life and viruses. All of them seem to be posttranscriptionally synthesized by protein enzymes (Boccaletto et al., 2018). RNA modifying enzymes; (i) work as (cofactor-dependent) stand-alone

proteins, (ii) are part of a protein complex that is required for modification or (iii) are guided by small nucleolar RNAs (snoRNAs).

Due to their diverse features, there is not a single way to classify RNA modifications; (i) universally distributed or confined to one domain, (ii) widespread localization or restricted to some RNA species, (iii) synthesized in a single- or multi-step enzymatic process, (iv) impacting the Watson-Crick interaction or not, (v) found on the nucleobase, the sugar and/or the 5'-phosphate, *etc.* (**Fig. 3A-E**).

The spatiotemporal regulation of RNA modifications is still an elusive aspect of RNA maturation. For example, snoRNA-dependent rRNA modifications (2'-O-methylation and pseudouridines, Fig. 3E) are usually synthesized in the nucleus, meaning early during the rRNA biogenesis (reviewed in (Sloan et al., 2017)). Nevertheless, rRNA modifications can also be exclusively cytoplasmic such as the yeast 18S-$m^{6,6}A_{1781}$ and $_{1782}$ modifications (Lafontaine et al., 1995). tRNAs are other noncoding RNAs for which the PTMs are closely related to the maturation process; methylation of position 40 (m^5C_{40}) was shown to be intron-dependent in the budding yeast $tRNA^{Phe}$, whereas m^1G_{37} requires intron removal to be installed. The RNA structure is impacted by the presence or absence of the intron, therefore changing the molecular identity that is essential for the enzyme recognition (Jiang et al., 1997). This example will be further discussed within the framework of this study but demonstrates that the decision *to modify or not to modify* is a tidy biological process.

B. Modified RNA species

Nucleic acids are not boring long polymers of only four types of nucleotides.

(Grosjean, 2009)

tRNAs are the most heavily modified RNA species with up to 25% and 15% of their ribonucleotides being modified in eukaryotes and bacteria, respectively (Grosjean, 2009). To date, it is estimated that one fifth of all known tRNA modifications is spread across all domains of life (Lorenz et al., 2017). Ribosomal RNAs are also widely modified – although to a lesser extent – with up to 2% of modified positions (Sloan et al., 2017). The abundance of rRNAs and tRNAs facilitated their study and the modification status of their building blocks. However, other RNA species have been known to carry PTMs for decades. Besides the well-characterized eukaryotic mRNA 5'-cap, internal modifications[2] are found in coding RNAs, such as the highly abundant m^6A which was detected in mRNAs more than forty years ago (Desrosiers et al., 1974). In bacteria, investigation about mRNA modifications is still in its early stages; a NAD-based cap-like structure has only been detected on bacterial RNA 5'-ends in 2009 (Fig. 3E) and the bacterial inosine[3] landscape is barely one year old (Bar-Yaacov et al., 2017; Chen et al., 2009).

[2] An internal modification refers to a residue that is found in the RNA body, in contrast with 5'- and 3'-end modifications.

[3] Inosine is a deaminated adenosine that is related to RNA editing. It will not be discussed in this manuscript.

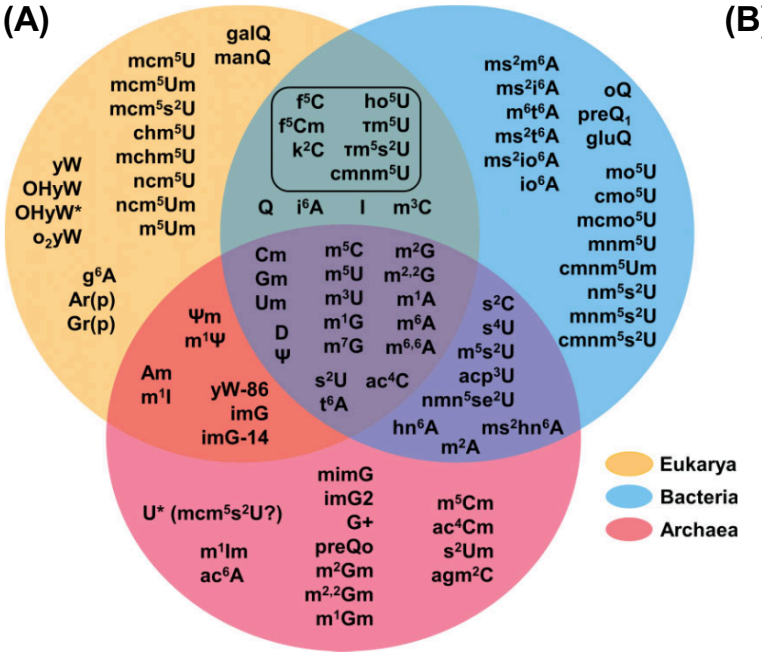
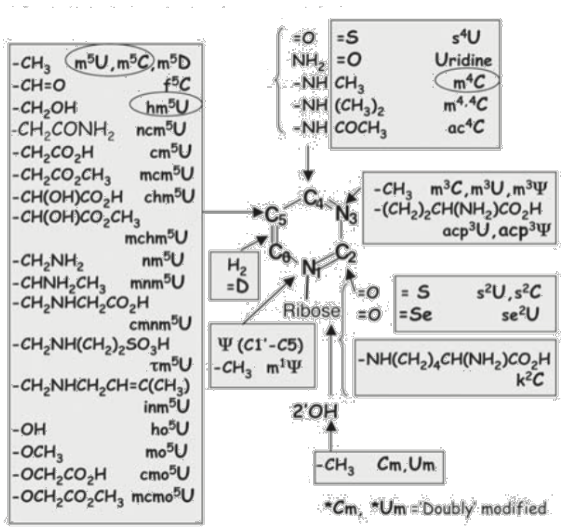
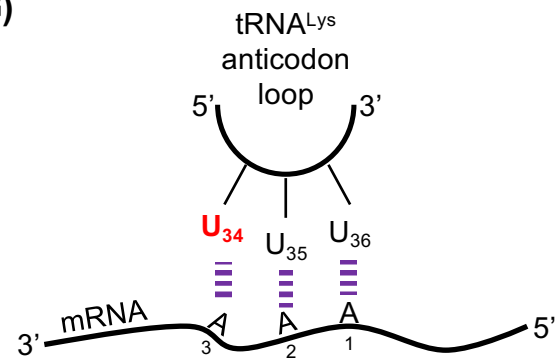

(A)
galQ
manQ
mcm5U
mcm5Um
mcm5s2U
chm5U
mchm5U
ncm5U
ncm5Um
m5Um
yW
OHyW
OHyW*
o2yW
g6A
Ar(p)
Gr(p)
f5C
ho5U
f5Cm
τm5U
k2C
τm5s2U
cmnm5U
Q
i6A
I
m3C
ms2m6A
ms2i6A
m6t6A
ms2t6A
ms2io6A
io6A
oQ
preQ1
gluQ
mo5U
cmo5U
mcmo5U
mnm5U
cmnm5Um
nm5s2U
mnm5s2U
cmnm5s2U
Cm
Gm
Um
D
Ψ
m5C
m5U
m3U
m1G
m7G
m2G
m2,2G
m1A
m6A
m6,6A
Ψm
m1Ψ
Am
m1I
yW-86
imG
imG-14
s2U
t6A
ac4C
s2C
s4U
m5s2U
acp3U
nmn5se2U
hn6A
ms2hn6A
m2A
U* (mcm5s2U?)
m1Im
ac6A
mimG
imG2
G+
preQo
m2Gm
m2,2Gm
m1Gm
m5Cm
ac4Cm
s2Um
agm2C
Eukarya
Bacteria
Archaea
(B)
m5s2U
tRNA only
m7G
widespread
localization
(C)
m6A
single modification
Q
hypermodification
(D)
methylation
(m7G)
Hoogsteen edge
Watson-Crick edge
methylation
(m1G)
Sugar edge
single/double
methylation
(m2G/m22G)
(E)
Ψ
nucleobase
modification
Gm
sugar
modification
m6Am
nucleobase and
sugar modifications
NAD
5'-end
modification

Figure 3 →

(F)

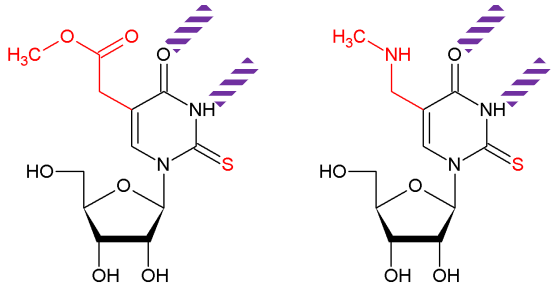


(G)



U_{34} (wobble)	
S. pombe	mcm^5s^2U
E. coli	mnm^5s^2U
mt-*S. cerevisiae*	$cmnm^5s^2U$

(H)



mcm^5s^2U **mnm^5s^2U**

→ **Figure 3 – Diversity of RNA modifications. (A)** Phylogenetic diversity of tRNA PTMs and distribution across the three domains of life. Framed modifications are eukaryotic organelle modifications (Jackman and Alfonzo, 2013). **(B)** m^5s^2U (5-methyl-2-thiouridine) is exclusively found on tRNAs (or at least has never been described elsewhere) whereas m^7G (7-methylguanosine) is found on different RNA species. **(C)** m^6A (N^6-methyladenosine) is synthesized in a single step process by a methyltransferase complex whereas the bacterial Q (queuosine) requires at least seven enzymatic reactions to be synthesized from G. **(D)** Examples of guanosine modifications that impair the Watson-Crick interaction (m^1G, m^2G, $m^{2,2}G$) or not (m^7G) (carbon in green, nitrogen in blue, oxygen in red and hydrogen in yellow) (Chawla et al., 2015). **(E)** RNA modifications occurring on the nucleobase (ψ, pseudouridine), on the ribose (Gm, 2'-O-methylguanosine), on both (m^6Am, N^6-2'-O-dimethyladenosine) or on the 5'-end (NAD, nicotinamide adenine dinucleotide). **(F)** Pyrimidine modifications (gray frames) found in RNA and DNA (circled) (Grosjean, 2009). **(G)** Schematic representation of the codon-anticodon Watson-Crick interaction (purple lines) between $tRNA^{LysUUU}$ and an AAA codon-containing mRNA. The wobble position of $tRNA^{LysUUU}$ is modified in mcm^5s^2U (5-methoxycarbonylmethyl-2-thiouridine) in the fission yeast (*S. pombe*), is modified in mnm^5s^2U (5-methylaminomethyl-2-thiouridine) in *E. coli* and is modified in mnm^5s^2U (5-carboxymethylaminomethyl-2-thiouridine) in the budding yeast mitochondrion (mt-*S. cerevisiae*). **(H)** Biochemical features of mcm^5s^2U and mnm^5s^2U. Hydrogen bonds involved in the Watson-Crick interactions are highlighted with purple lines (Boccaletto et al., 2018).

In Eukarya, snoRNAs and snRNAs carry 5'-end and internal modifications but have a narrow range of PTMs compared to the tRNAs and rRNAs (Boccaletto et al., 2018; Cantara et al., 2011). Other reports also revealed the presence of modified nucleotides at various transcriptomic positions; primary miRNAs require m^6A (Fig. 3C) to be properly processed by the RNase DROSHA (Alarcon et al., 2015); the human telomere RNA stability is increased by the pseudouridylation (Fig. 3E) of one of its stems, which was shown to boost the processivity of the telomerase itself *in vitro* (Kim et al., 2010); the methylation status of the XIST lncRNA (*X-inactive specific transcript*; modulating the silencing of one X chromosome during the female mammalian development) was experimentally correlated to its ability to repress transcription (Patil et al., 2016); *etc.*

In Bacteria, the RNA modifications world is less extensively described but the non-canonical transfer-messenger RNAs (tmRNAs) were shown to hold at least two different modifications – ψ (pseudouridine) and m^5U (5-methyluridine) (Felden et al., 1998). A tmRNA is characterized both by a tRNA-like domain (TLD) with a D-loop-like conformation at its 5'-end and by a sequence coding for a proteolytic cleavage-inducing tag. This latter is added to a nascent polypeptide when the ribosome is stalled in the so-called *trans*-translation process and therefore causes the degradation of the truncated protein (reviewed in (Mace and Gillet, 2016)). The possible crosstalk between the tmRNA modification pattern – especially on the TLD – and its ability to recognize stalled ribosomes is still unknown.

RNA species are structurally and functionally different but they all seem to be targeted by modifying enzymes that install a large set of RNA modifications. Although it is a tedious work, the functionality of PTMs has been partially unraveled and examples are given below.

C. Functionality of RNA modifications

Understanding the scope and mechanisms of […] dynamic RNA modifications, which could be termed 'RNA epigenetics', represents a new frontier in research for chemical biologists.

(He, 2010)

Why is there *that* modification on *that* RNA? This general question is of high interest in RNA biology because the fate and functionality of an RNA also rely on the modifications profile. For brevity, only four functional examples will be shortly described to give an insight about how a posttranscriptional modification can be meaningful for RNA.

i. Methylation of a lncRNA in Eukarya

The functionality of a noncoding RNA was illustrated hereabove with the MALAT1 example (see Introduction I.B.iv.a); this long noncoding RNA localizes in the nuclear speckles and is associated with pre-mRNA splicing. In this section, details are given about how this functionality is acquired through an RNA modification. MALAT1 was indeed shown to be regulated by the m^6A-switch mechanism. When the MALAT1 adenosine 2577 is methylated, RNA pull-down assays demonstrated an enhanced

affinity for the RNA-binding protein (RBP) HNRNPC. In other words, the MALAT1-m^6A_{2577} is more strongly bound to HNRNPC than its MALAT1-A_{2577} counterpart. The same study highlighted that the m^6A-methyltransferase KO or *HNRNPC* KO led to the dysregulation of mRNA abundance and alternative splicing (Liu et al., 2015). Because MALAT1 is considered as a pre-mRNA processing element (Wilusz et al., 2008), it is tempting to establish a direct link between MALAT1 methylation and alternative splicing regulation but experimental evidence are still lacking. However, the modulation of an RNA-protein interaction through a PTM reveals how an RNA modification acquires functionality.

ii. A rare RNA modification in extreme thermophilic Bacteria

Extreme thermophilic bacteria optimally grow above 60°C. High temperature leads to tRNA destabilization, therefore forcing the extreme thermophilic organisms to acquire molecular specificities. An example of such an adaptation relies on the biosynthesis of m^5s^2U (thiolated derivative of a methylated uridine, Fig. 3B) at position 54 in almost all tRNAs of *Thermus thermophilus* (whereas absent in non-thermophilic bacteria). The thiolation reaction leading to m^5s^2U formation is dependent of the temperature (minimum ~ 55°C), while the archaeon *P. furiosus* accumulates m^5s^2U at high temperatures (Kowalak et al., 1994; Shigi et al., 2006). With this modification, the tRNA thermal stability is then ensured and allows proper translation (Yokoyama et al., 1987). Coupling RNA modification to environmental cues is therefore essential for high-temperature survival.

iii. Methylation-dependent hijacking by viral RNAs

m^6A is a highly dynamic modification and is usually introduced with the concepts of *writer*, *reader* and *eraser*, that is to say enzymes that *write*, *read* or *erase* the modification. m^6A is synthesized by methyltransferases (= *writers*), interacts with specific RBPs (= *readers*) and is *wiped out* by demethylases (= *erasers*).

As viruses rapidly evolve, they acquire molecular specificities to hijack the host cell machinery. RNA modifications-related enzymes are no exception. In 2016, Cullen and colleagues mapped m^6A on HIV-1 mRNAs and highlighted a cluster of modification on the 3'-UTR. They further showed that the loss of m^6A deposition led to reduced viral RNA expression and that the overexpression of m^6A readers (YTHDF proteins) resulted in an increase of viral proteins production (Kennedy et al., 2016). This study strongly suggests that; (i) the viral genome has evolved to express mRNAs containing m^6A consensus sequences that are recognized by the m^6A readers, (ii) the m^6A-containing mRNAs are bound by the YTHDF proteins that promote viral RNA translation and (iii) HIV uses the beneficial m^6A dynamic modification to optimize its replication.

iv. The wobble modification: a universal role

The modifications that occur near or on the anticodon are of high importance for the stabilization of the codon-anticodon interaction and subsequent optimal translation. The third position of any codon interacts with the 34^{th} position of its cognate tRNA, the so-called wobble position. This nucleotide is frequently modified and decorated by a plethora of different RNA modifications (Boccaletto et al., 2018), notably by pyrimidine

derivatives (**Fig. 3F**). An example is the hypermodification of the $tRNA^{LysUUU}$ wobble position (**Fig. 3G**). The 5'-nucleotide of the UUU anticodon is heavily modified with an mcm^5s^2U modification in the cytoplasmic $tRNA^{Lys}$ of Eukarya and with a closely related mnm^5s^2U modification in bacterial counterparts (**Fig. 3H**). It should be noticed that mnm^5s^2U modification is not exclusively found in Bacteria. Remarkably, the $tRNA^{LysUUU}$ wobble modification was shown to be crucial for translational fidelity and optimization both in prokaryotes and eukaryotes (Schaffrath and Leidel, 2017). For example, loss of mcm^5s^2U in the fission yeast is not lethal but results in cell cycle defects. At the molecular level, the phenotype was partly explained by the translational control of an mRNA coding for a cell cycle regulator; this mRNA was enriched in AAA codons, explaining why its expression was $tRNA^{Lysmcm5s2UUU}$-dependent. Strikingly, the engineered exchange of the AAA codons to AAG codons was sufficient to rescue the WT level of protein production in a mutant lacking the mcm^5s^2U modification (Bauer et al., 2012). Similarly, the reading frame maintenance was shown to require the mnm^5s^2U-modified $tRNA^{LysUUU}$ in bacteria (Bregeon et al., 2001; Urbonavicius et al., 2001). The phylogenetic proof-of-concept also relies on the regulation of mitochondrial translation; the budding yeast mitochondrial $tRNA^{Lyscmnm5s2UUU}$ (encoded by the mitochondrial genome) is sufficient to decode AAA and AAG codons in normal conditions. However, the wobble position is hypomodified at elevated temperatures and the mitochondrial translation machinery requires the import of the cytoplasmic $tRNA^{LysCUU}$ to prevent translational deficiency, demonstrating that the proper modification of the $tRNA^{LysUUU}$ wobble position is critical for optimized protein production (Kamenski et al., 2007).

The functional importance of the wobble position as briefly described here shows how a tRNA modification fine-tunes the translation, which is a more direct effect compared to what is observed with PTMs that are found on the tRNA body.

III. FOCUS ON DIHYDROURIDINE

A. Biochemical and structural identities

Dihydrouridine is structurally unique […].

(Davis, 1998)

The dihydrouridine (D) RNA modification is a modified pyrimidine nucleoside whose corresponding nucleobase is 5,6-dihydrouracil. D is synthesized from uridine (U) by hydrogenation (Green and Cohen, 1957) (Fig. 3F, **Fig. 4A**). Reduction of the uridine C5-C6 bond generates a saturated nonplanar and nonaromatic nucleobase that is a landmark of dihydrouridine (**Fig. 4B**).

Although the chemical synthesis of dihydrouracil was already reported in 1896, its first detection in a biological sample dates back from 1952 when it was isolated from the beef spleen (Ehrlich et al., 1952). Using *in vitro* approaches, dihydrouridine monophosphate (dihydro-UMP) was shown to be efficiently introduced into RNA molecules, but whether D was a normal component of the RNA was still to be determined (Carr and Grisolia, 1964; Royburman et al., 1965). D was then reported as

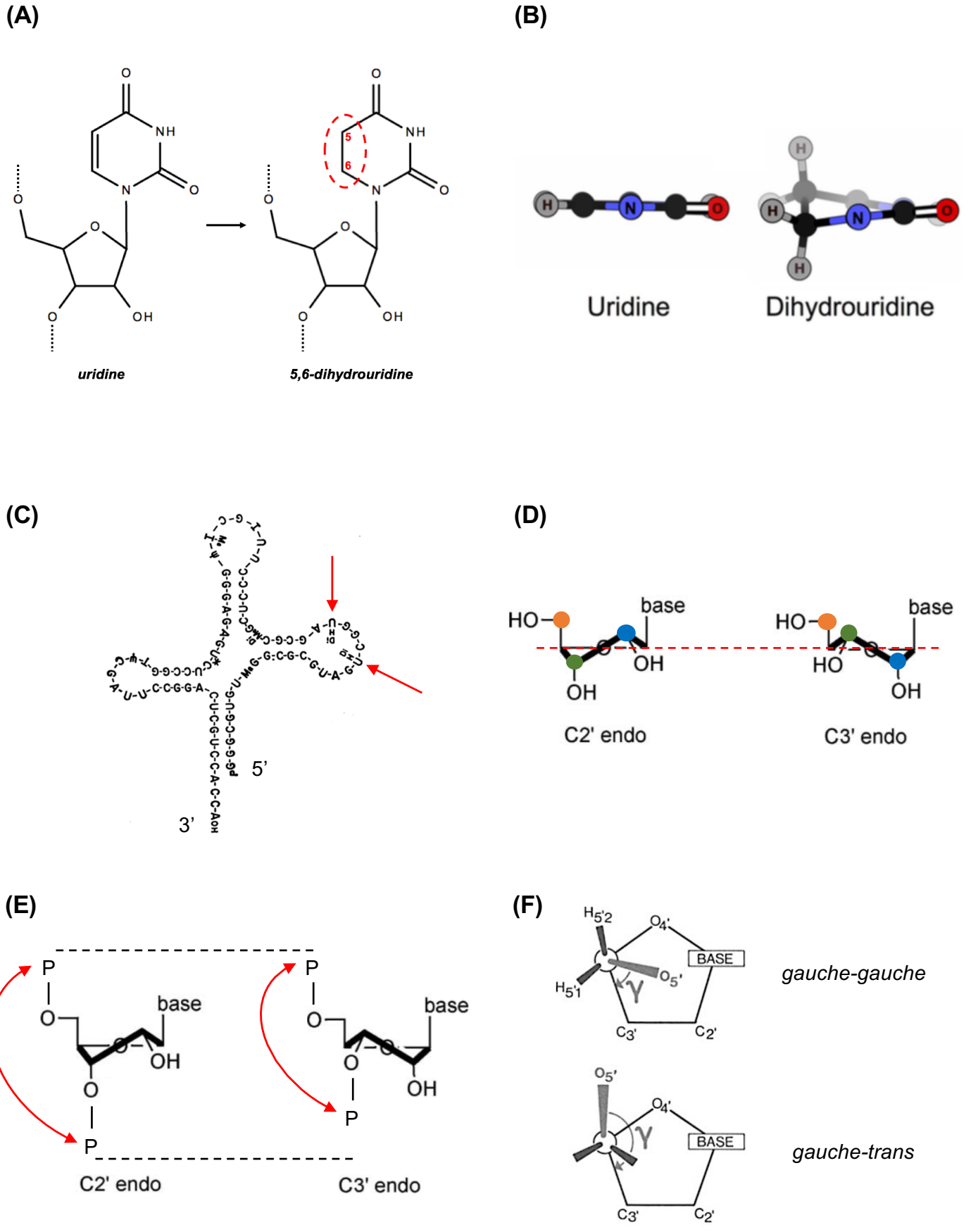


Figure 4 – Biochemistry of dihydrouridine. (A) Reduction of uridine into dihydrouridine. **(B)** Dihydrouracil is a nonplanar nucleobase (carbon in black, nitrogen in blue, oxygen in red and hydrogen in gray) (Whelan et al., 2015). **(C)** First published structure of a ribonucleic acid (yeast $tRNA^{Ala}$) where D (red arrows) are shown in a loop at the 5'-end (Holley et al., 1965). **(D)** Schematic representation of ribose pucker. C5' (orange dot) is considered as being *above* the C4'-O'-C1' plane (red dashed line). **On the left,** C2*'-endo* has the C2' (blue dot) above the plane. **On the right,** C3*'-endo* has the C3' (green dot) above the plane (Moreau et al., 2011). **(E)** C2'-*endo* pucker produces a longer 5'-phosphate/3'-phosphate distance and therefore spans the polynucleotide (Moreau et al., 2011). **(F)** Schematic representation of ribose *gauche-gauche* and *gauche-trans* conformations. C5' (white dot) adopts different torsion angles (γ) that modulate the positionning of C5'-bound atoms ($H_{5'1}$, $H_{5'2}$ and $O_{5'}$) (Saenger, 1984).

a naturally occurring component of yeast $tRNA^{Ala}$ and even included in the first published structure of a ribonucleic acid (Holley et al., 1965; Madison and Holley, 1965) (**Fig. 4C**). Despite the lack of evidence, Visser and colleagues were already discussing the possibility that "non-random distribution of hydrogenated pyrimidine [dihydrouridine] may be explained more readily by a process of enzymatic hydrogenation at the polynucleotide level" ((Royburman et al., 1965), p. 297).

In the next decades, optical studies along with X-ray and NMR crystallographic analyses were the main focus to elucidate the biochemical properties of dihydrouridine. The structures of the nucleobase, the nucleotide, of D-containing tRNAs and D-containing oligoribonucleotides led to the conclusions that; (i) the carbon 6 of the nucleobase (C6) is out of the plane after the C5-C6 double bond reduction, (ii) the deviation from the planar nature of the pyrimidine results in the loss of the stacking ability with neighboring nucleobases, (iii) the C2'-*endo* conformation is adopted by the ribose moiety whereas RNA prefers the C3'-*endo* conformation (**Fig. 4D**) and (iv) the C2'-*endo* pucker is propagated to the 5'-nucleotide (references and comments in **Table 1**). In other words, the complete destacking of the bases and the unusual adoption of C2'-*endo* ribose pucker make dihydrouridine a unique modification (Davis, 1998). The structural properties of dihydrouridine include the potential destabilization of the RNA structure (by promoting C2'-*endo* conformation) and molecular flexibility (by spanning the sugar-phosphate backbone, **Fig. 4E**). In parallel, the D crystal structures revealed the adoption of *gauche-trans* or *trans-gauche* conformations around the C4'-C5' ribose bond, rather than the common *gauche-gauche* rotamer (**Fig. 4F**, references in Table 1). Finally, the dihydrouracil nucleobase was shown to be in *anti* orientation in respect to the ribose moiety (i.e. with C2=O pointing away from the sugar) (references in Table 1).

B. Methods of detection

Over the decades, multiple experimental techniques were developed for the identification and localization of RNA modifications. Most of them are quite laborious […].

(Motorin et al., 2007)

i. Sodium borohydride versus alkali

Distinguishing the RNA modification from its canonical nucleotide is challenging. Investigations on chemical reactions specifically acting on D started before its description as a natural PTM (**Table 2**). The dihydrouridine undergoes ring opening upon sodium borohydride ($NaBH_4$) or alkaline (OH^-) treatments, resulting in the formation of an ureido-group (NH_2CONH) linked to an alcohol or a carboxylic acid, respectively (**Fig. 5**) (Batt et al., 1954; Cerutti and Miller, 1967; Igo-Kemenes and Zachau, 1969). The accumulation of ureido-groups can be quantified by a colorimetric assay (Jacobson and Hedgcoth, 1970). The ribosylureidopropanol (D + $NaBH_4$) is either used for labeling of RNA with a fluorescent dye (point iii.) or for the cleavage of ureidopropanol upon acid conditions (H^+) (Betteridge et al., 2007; Cerutti et al., 1968a; Cerutti et al., 1968b; Cohn and Doherty, 1955; Pan et al., 2009; Wintermeyer et al., 1979; Wintermeyer and Zachau, 1971, 1974). The ribosylureidopropionic acid (D + OH^-) is used as a semi-quantitative tool following the breakdown between ribose and

structural and optical studies	comments	references
dihydrouracil	C5-C6 and N1 out of the base plane expectation for Watson-Crick interaction impairment	(Rohrer and Sundaralingam, 1970)
dihydrouridine	deviation from planarity *anti* orientation C2'-*endo* conformation	(Sundaralingam et al., 1971a)
	anti orientation favoured *gauche-trans* and *trans-gauche* conformations	(Sundaralingam et al., 1971b)
	deviation from planarity *anti* orientation preference for C3'-*exo* and C2'-*endo* conformations preference for *gauche-gauche* conformation	(Deslauriers et al., 1971)
	deviation from planarity *anti* orientation C2'-*endo* conformation favoured *gauche-trans* and *trans-gauche* conformations	(Suck et al., 1971)
	C6 out of the base plane no base stacking and *anti* orientation C2'-*endo* conformation favoured *gauche-trans* and *trans-gauche* conformations	(Suck et al., 1972)
dinucleoside phosphate: DpA, ApD, GpD	DpA stacking ApD and GpD low stacking	(Formoso et al., 1971)
yeast or *E. coli* tRNAs	formation of a D-loop with increased hydrophobicity D provides extra flexibility interaction of D-loop and TψC-loop	(Davis et al., 1986; Jack et al., 1976; Quigley and Rich, 1976; Westhof et al., 1985; Westhof and Sundaralingam, 1986; Woo et al., 1980)
trinucleoside phosphate: Dp(acp^3U)pA	C2'-*endo* conformation and upstream propagation	(Stuart et al., 1996)
trinucleoside phosphate: ApDpA	C6 out of the base plane no base stacking enhanced C2'-*endo* conformation at low temperature (5°C) and downstream propagation	(Dalluge et al., 1996b; Deb et al., 2014)

Table 1 – Seminal studies on optical and structural properties of dihydro-uracil/-uridine. The table is divided into three columns; (i) studied chemical entities, (ii) main conclusions relative to the nucleobase (written in green) or to the ribose (written in blue) and (iii) references. C (carbon), N (nitrogen), A (adenosine), D (dihydrouridine), G (guanosine), p (5'-3' phosphodiester bond), acp^3U (3-(3-amino-3-carboxypropyl)uridine).

principles	tRNA	comments	references
alkaline hydrolysis	/	sodium hydroxide → cleavage at the N3-C4 linkage of dihydrouracil	(Batt et al., 1954)
β-alanine detection	Y	sodium hydroxide → D-ring opening → partial β-alanine formation → ninhydrin colorimectric assay	(Magrath and Shaw, 1967)
sodium borohydride reduction	Y	sodium borohydride → cleavage at the N3-C4 linkage of dihydrouridine	(Cerutti et al., 1967)
hydrochloric acid hydrolysis	B, Y, M	sodium borohydride → D-ring opening → hydrochloric acid → cleavage of the glycosodic bond	(Cerutti et al., 1968a and b; Cohn and Doherty, 1955)
loss of absorbance	Y	loss of absorbance at 265nm upon mild sodium hydroxide treatment	(Molinaro et al., 1968)
cleavage at D position	Y	sodium borohydride → D-ring opening → aniline treatment → RNA chain cleavage	(Beltchev and Grunberg-Manago, 1970)
ureido-group detection	B, Y, M	sodium hydroxide → D-ring opening → solution neutralization → colorimetric assay after iron chloride addition (550nm)	(Jacobson and Hedgcoth, 1970)
replacement of D by proflavine or EtBr	Y	sodium borohydride → D-ring opening → incubation with the dye	(Wintermeyer and Zachau, 1971)
microarray	Y	differential fluorescent labeling of tRNA from WT and *dus* strains → annealing on a microchip → quantification of A-U *vs* A-D interaction	(Peng et al., 2003)
primer extension	B, Y	sodium hydroxide → D-ring opening → assessment of RT termination by primer extension	(Betteridge et al., 2007; Xing et al., 2004)
fluorescent labeling with rhodamine 110	B	*in vitro* tRNA → *in vitro* dihydrouridylation → sodium borohydride → incubation with the dye	(Betteridge et al., 2007)
replacement of D by Cy3 or 5	B, Y	sodium borohydride → D-ring opening → incubation with the dye	(Pan et al., 2009)
benzoyhydrazide addition	B	sodium borohydride → THU formation → incubation with benzohydrazide	(Kaur et al., 2011)

Table 2 – Chemical reactions and techniques specifically applicable to D. The table is divided into four columns; (i) list of principles, (ii) phylogenetic origin of the studied tRNA; B (bacterial), Y (yeast), M (mammalian), (iii) general comments on the principle. EtBr (ethidium bromide), N (nitrogen), C (carbon), RT (reverse transcritpion), WT (wild type), *dus* (dihydrouridine synthase gene), A-U (adenine-uridine Watson-Crick interaction), A-D (adenine-dihydrouridine Watson-Crick interaction), Cy3/5 (cyanines 3 or 5), THU (tetrahydrouridine) and (iv) references (only the seminal works are cited although most of the techniques were implemented in other studies).

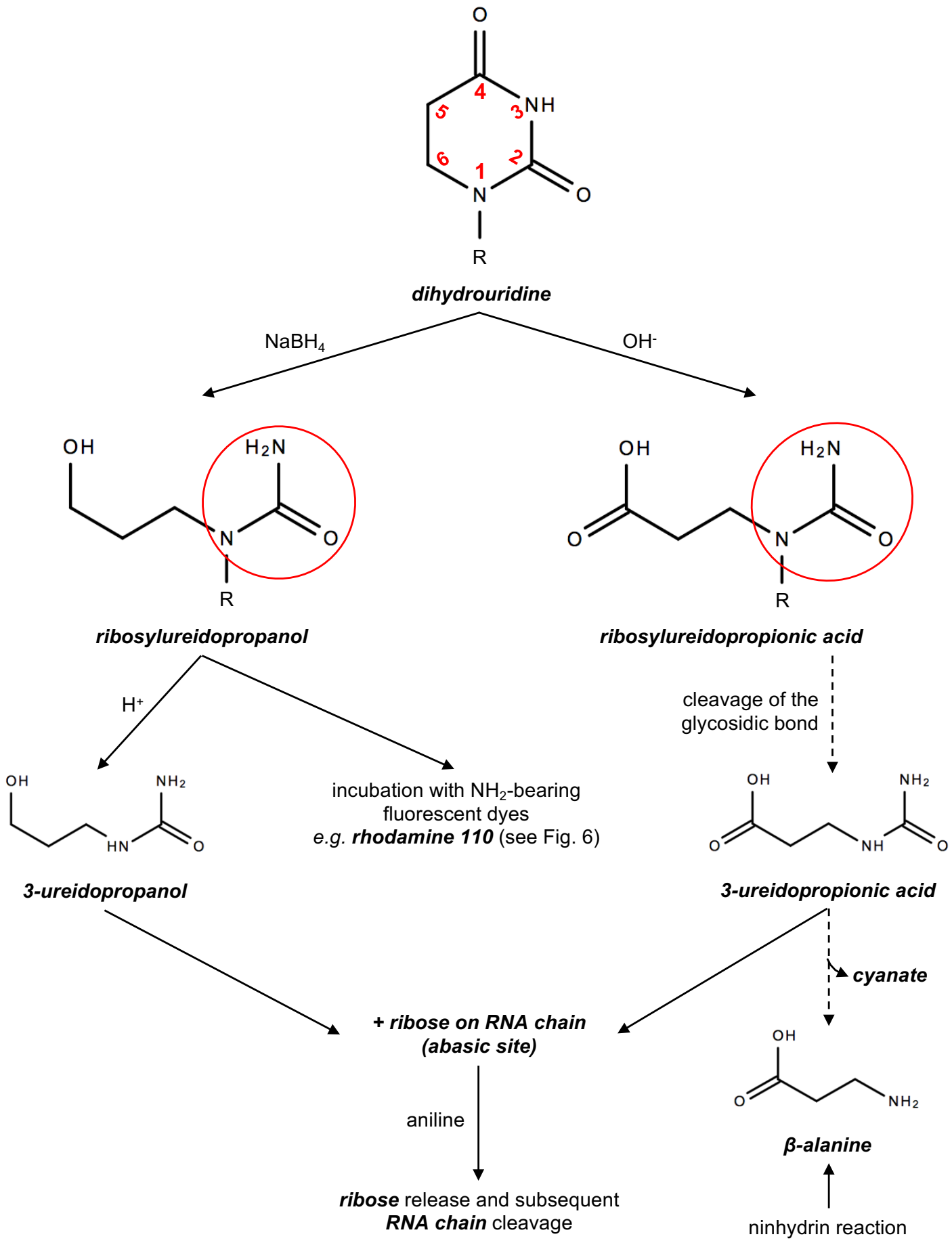


Figure 5 – Molecular fate of D upon sodium borohydride ($NaBH_4$) or alkaline (OH^-) treatments. R stands for ribose attached to the RNA chain. Red circles highlight ureido-groups. H^+ stands for acid conditions. Details in the text.

ureidopropionic acid, this latter being decomposed to β-alanine that serves as a substrate for a colorimetric assay with ninhydrin (House and Miller, 1996; Magrath and Shaw, 1967). The D-ring disruption upon OH^- condition was also shown to generate an RT (reverse transcription) termination assessed by primer extension (**Fig. S2**) (Xing et al., 2004). In both cases (D + $NaBH_4$ or OH^-), it is possible to get an abasic site (nucleobase-free ribose) that leads to the cleavage of RNA chain with aniline treatment (Beltchev and Grunberg-Manago, 1970; Wintermeyer and Zachau, 1970).

ii. D and Watson-Crick interactions

The comparison of binding properties of polyU- *vs* polyD-nucleotides with polyA revealed a diminished ability of polyD to interact with polyA although the reduction occurs at Hoogsteen edge (C5=C6) of the nucleobase whereas the hydrogen bonds are formed from positions 3 and 4 (N3-C4=O) (Cerutti et al., 1966). The decreased binding capacity is nevertheless too weak to induce a clear RT termination in a primer extension context (Motorin et al., 2007). Phizicky and colleagues used this low binding feature between D and A for the implementation of a microarray-based technology in order to assess the presence or absence of a D residue at a specific tRNA position in yeast mutants (see point C.i.b.) (Xing et al., 2004). In brief, total RNAs are extracted from WT (wild type) and D-defective mutant strains and are differentially labeled with fluorescent dyes emitting at non-overlapping wavelengths (*e.g.* WT RNA is labeled with a dye emitting at 546nm and the mutant RNA is labeled with a dye emitting at 647nm). The capacity of fluorescent RNAs to hybridize with tDNA probes (DNA sequences of a tRNA) coated on a chip is dependent on the sequence itself; carrying a D ("low" signal due to weak A-D interaction) or not ("high" signal due to strong A-U interaction) (Peng et al., 2003).

iii. D fluorescent labeling

As stated above, the reduction of D by sodium borohydride is a prerequisite for the subsequent incorporation of fluorescent molecules, and more specifically of NH_2-dyes (*e.g.* proflavine, rhodamine, cyanine hydrazide). Figure 5 shows that the D-ring opening forms ribosylureidopropanol in a $NaBH_4$-dependent manner. However, this commonly accepted N3-C4 cleavage has been recently challenged (Kaur et al., 2011). Cooperman and colleagues induced tRNA reduction with sodium borohydride and performed labeling with NH_2-containing fluorophores nucleophiles. By combining TLC (thin layer chromatography) and mass spectrometry, they detected tetrahydrouridine (THU) instead of the expected ribosylureidopropanol. Based on their results, they proposed that the dihydrouridine C4 carbonyl group is reduced by the H^- donor $NaBH_4$ to THU. Upon the addition of an NH_2-dye in acid conditions, a nucleophilic substitution occurs on the C4 hydroxyl group by formation of a Schiff base-bearing intermediate called tetrahydrocytidine (THC) and consecutive fluorophore binding. **Fig. 6** summarizes both mechanisms in the context of the addition of the rhodamine 110 fluorophore.

iv. General methods

Besides the above detailed methods that are selective for dihydrouridine only (along with a few other RNA modifications sharing the same chemical reactivity, see Table 3 in Discussion and Perspectives), general methods are also used to detect and possibly quantify D.

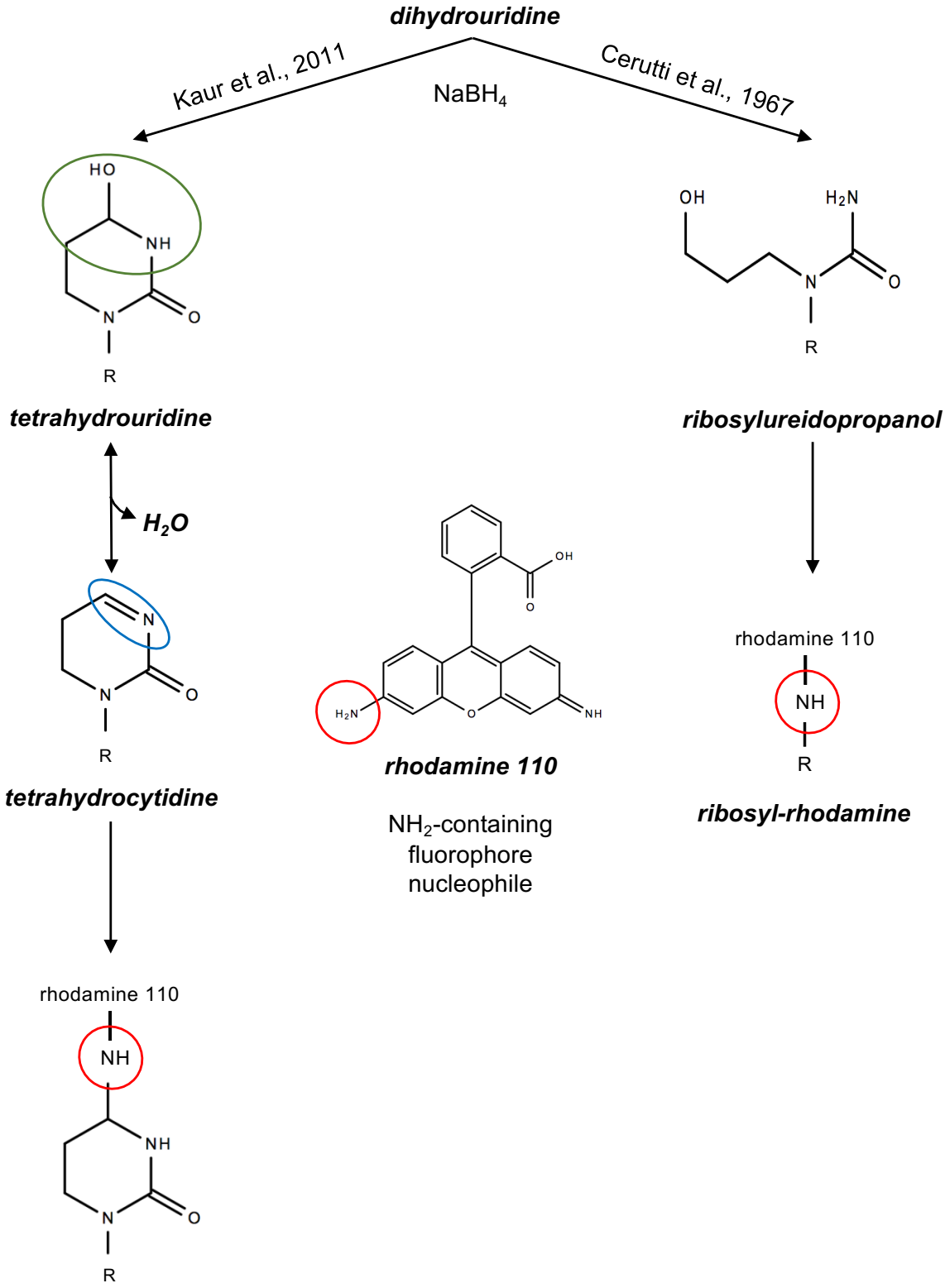


Figure 6 – Proposed mechanism of D rhodamine labeling following sodium borohydride ($NaBH_4$) reduction. On the left, dihydrouridine is reduced to tetrahydrouridine that is characterized by an electrophilic carbonilamine (green circle), including the C4 hydroxyl group (C-OH). Upon addition of the nucleophilic rhodamine 110 in acid conditions, the tetrahydrocytidine intermediate is formed with its reactive Schiff base (bleu circle). Covalent binding of rhodamine 110 occurs by substitution of the C4 hydroxyl group. **On the right**, ureidopropanol generated by D-ring opening is replaced by rhodamine 110. R stands for ribose attached to the RNA chain.

TLC is a long-standing technique that chemically separates radioactively-labeled (modified) nucleotides. D was detected *in vivo* after addition of ^{14}C- and ^{3}H-uracil in the growth medium, followed by RNA extraction, digestion and separation by bi-dimensional TLC (Jacobson and Hedgcoth, 1970). One-dimension TLC was efficiently used to assess the formation of D on an *in vitro* transcribed tRNA incubated with GST-purified putative dihydrouridine synthases and various cofactors (Xing et al., 2002). The detection was also accomplished by ^{3}H-labeling of the ribose moiety. Briefly, the RNA was enzymatically digested, oxidized with an inorganic salt – leading to the opening of the ribose ring –, and reduced by ^{3}H-borohydride that *attacks* the opened ribose ring by addition of ^{3}H. Labeled mononucleotides are then spotted on a cellulose membrane for separation in two dimensions (Randerath et al., 1980).

Reverse-phase high performance liquid chromatography (HPLC) is commonly used to sort the nucleosides according to their biochemical properties and are then detected by UV. The dihydrouridine has however no appreciable absorbance in the UV spectrum. Moreover, D is the earliest eluting nucleoside and is therefore sometimes barely distinguished from contaminants with short retention times. These technical issues are nevertheless overcome by the detection at 254nm (Buck et al., 1983; Pomerantz and McCloskey, 1990) and improvement of D detection by UV is achieved by lowering the wavelength at 210 or 230nm (Gehrke and Kuo, 1990; Topp et al., 1993; Xing et al., 2004; Xing et al., 2002). Accurate determination of D is readily obtained by liquid chromatography coupled to mass spectrometry (LC-MS) (Dalluge et al., 1996a; Edmonds et al., 1985).

C. The biology of dihydrouridine

Our understanding of the physiological role of dihydrouridine is still in its infancy […].

(Rider et al., 2009)

Dihydrouridine is found in RNA from all domains of life but is mostly described within the framework of tRNA biology because it is essentially found on them. Bujnicki and colleagues aligned 602 tRNA sequences from more than 100 prokaryotic and eukaryotic species. Distribution of PTMs revealed that dihydrouridine is the second most prevalent tRNA modification (925 counts) after pseudouridine (1,164 counts) (Machnicka et al., 2014). The dihydrouridylated positions include the canonical D_{16}, D_{17}, D_{20}, D_{20a}, D_{20b} and D_{47} and the rare noncanonical D_{14}, D_{17a}, D_{21} and D_{48}. Strikingly, among the most frequent modified positions, D_{20} is the third with 323 occurrences and D_{16} is the fifth with 216 occurrences. These highly conserved D are positioned in the so-called *D-loop* – for which it is named – that has a pivotal role in the establishment of secondary and tertiary structures of tRNAs (**Fig. S3**) (Dyubankova et al., 2015; Grosjean et al., 1998). The biochemical specificities of D (see Introduction III.A.) play a role in the cloverleaf-related tRNA secondary structure and in the L-shaped tRNA tertiary structure that is achieved through D- and T-loops interaction (reviewed in (Vare et al., 2017)). Although the most important residues for the kissing D/T loops are not dihydrouridines, a compilation of crystal structures highlighted a set of base pairing events where D is involved through various types of interactions (*cis* or *trans* interactions between Watson-Crick, Hoogsteen and sugar edges) (Chawla et al., 2015;

Seelam et al., 2017). In other words, D is a highly conserved and prevalent posttranscriptional modification that is also a key player for tRNA structure.

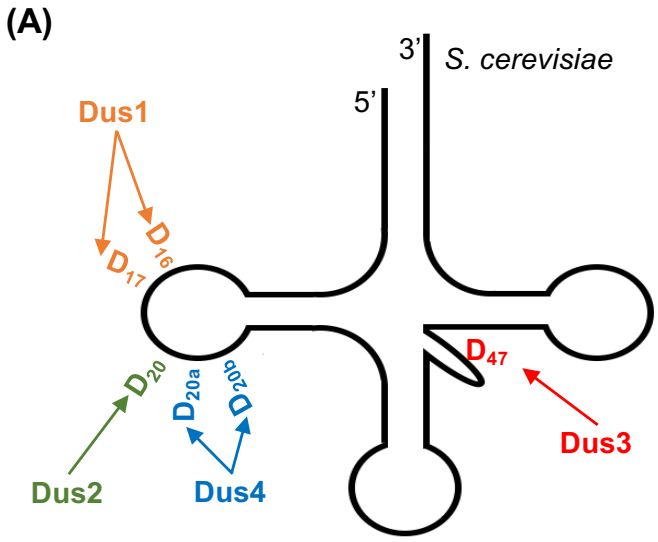
i. Specificities in Eukarya

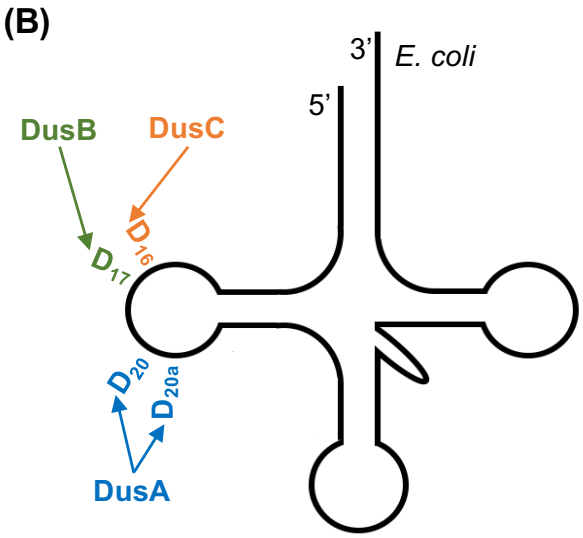
a. Generalities

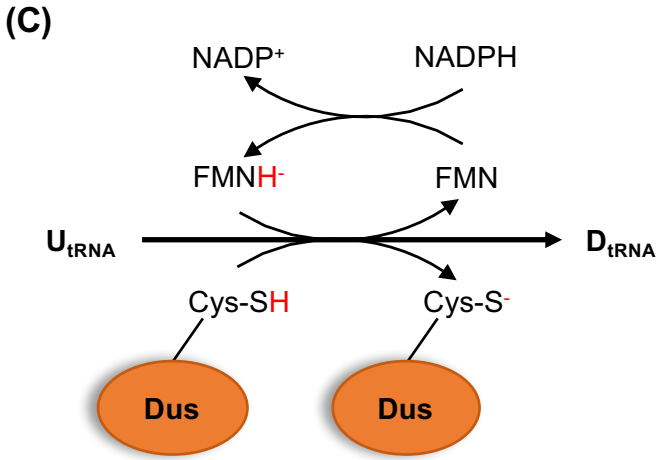
Historically, the dihydrouridine was associated with plant and mammalian histone-bound RNAs (Huang and Bonner, 1965; Shih and Bonner, 1969). Later, D was detected on an enzymatic digest of rat U5 snRNA, but this observation was never confirmed since then (Krol et al., 1981). Well-described eukaryotic dihydrouridylated positions are found on tRNAs; all tRNA species have been described with at least one dihydrouridine, with the exception of $tRNA^{selenocysteine}$. Particularly, none of the eighteen described cytoplasmic $tRNA_i^{Met}$ have any D_{17}, D_{20}, D_{20a}, or D_{20b} and only 28% of them have a D_{16} (against 76% for the 210 others cytosolic tRNAs). In mitochondrial tRNAs (mt-tRNAs), all six canonical D are found, except D_{20b} (Cantara et al., 2011; Machnicka et al., 2014). Importantly, there are dihydrouridylated mt-tRNAs that are encoded by the mitochondrial genome; at least three mammalian mt-tRNAs are known to have a D_{20} and accordingly, the putative cognate enzyme (see point b.) was shown to localize in human and murine mitochondria (Calvo et al., 2016; Suzuki and Suzuki, 2014). It has been shown that a set of mammalian mt-tRNAs have a truncated cloverleaf structure by lacking the D-loop (Arcari and Brownlee, 1980; de Bruijn et al., 1980). Surprisingly, these tRNAs seem to adopt a functional tertiary structure by establishing unique interactions in a Mg^{2+}-dependent manner (de Bruijn and Klug, 1983; Jones et al., 2006).

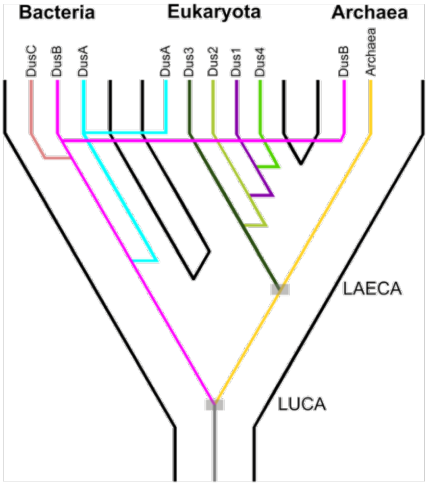
b. Eukaryotic Dus enzymes

Determination of RNA modifying enzymes has lagged behind since the discovery of the first RNA modification (Davis and Allen, 1957; El Yacoubi et al., 2012). A budding yeast strain – characterized by the production of multiple isoaccepting tRNAs in specific growth conditions – was shown to have a reduced amount of D (Lo et al., 1982). However, the genetic basis of this peculiarity is still unresolved and therefore did not allow to isolate a potential *dihydrouridine synthase* candidate (*dus* or Dus). Phizicky and colleagues accomplished this work in 2002 – 37 years after D was shown to be a natural RNA modification – in *S. cerevisiae* (Xing et al., 2002). More specifically, they performed a Dus activity assay by combining *in vitro* transcribed tRNAs (^{32}P-labeled at U positions) and a collection GST-ORF fusion proteins. D formation was assessed by TLC. Dissection of the positive candidates led to the isolation of a unique ORF that was called *dus1*. Furthermore, the *in silico* alignment of the Dus1 sequence revealed that three other similar proteins were encoded in the yeast genome, and that the Dus1 homologs were found in *S. pombe*, *C. elegans*, *D. melanogaster*, *A. thaliana*, *H. sapiens* and *E. coli*. In parallel, the detailed investigation on Dus activity highlighted that D was synthesized on $tRNA^{Phe}$ incubated with Dus1 or on $tRNA^{Leu}$ with Dus2, but not conversely. They therefore hypothesized that the Dus proteins could be substrate-specific enzymes. The high specificity of Dus1, 2, 3 and 4 was demonstrated in a microarray-based experiment (see Introduction III.B.ii and Table 2) for yeast cytoplasmic tRNAs; Dus1 targets U_{16} and U_{17}, Dus2 targets U_{20}, Dus4 targets U_{20a} and U_{20b} and Dus3 targets U_{47} (**Fig. 7A**). Finally, the budding yeast quadruple *dus* mutant was shown to be viable and to lack any detectable dihydrouridine (Xing et al., 2004).









(E)

enzyme	*Sp.*	dm	references
Dus	*T. maritima*	B	(Park et al., 2004)
DusA	*T. thermophilus*	B	(Yu et al., 2011)
hDUS2	*H. sapiens*	E	(Griffiths et al., 2012)
DusC	*E. coli*	B	(Chen et al., 2013)
DusC	*E. coli*	B	(Byrne et al., 2015)
hDUS2	*H. sapiens*	E	(Whelan et al., 2015)
hDUS2	*H. sapiens*	E	(Bou-Nader et al., 2015)
DusB	*E. coli*	B	(Bou-Nader et al., 2018)

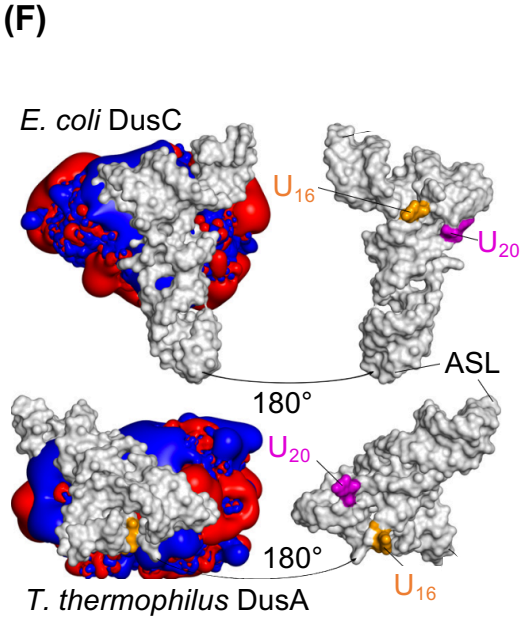


Figure 7 – The biology of dihydrouridine. (A) Specificity of yeast Dus enzymes for cytoplasmis tRNAs. **(B)** Specificity of *E. coli* Dus enzymes for tRNAs. **(C)** Putative enzymatic mechanism of Dus enzymes. Details in the text. **(D)** Proposed evolutive scenario of Dus enzymes across the three domains of life. LUCA (last universal common ancestor). LAECA (last archaeal and eukaryotic common ancestor) (Kasprzak et al., 2012). **(E)** Currently available structures of Dus enzymes in different species (*Sp.*) from different domains (dm; B for Bacteria and E for Eukarya). The references are indicated in the last column. **(F)** Bacterial Dus enzymes bind tRNA in reverse orientations. DusC modifies U_{16} and DusA modifies U_{20}. ASL (anticodon stem and loop domain) (Byrne et al., 2015).

c. D and cancer

The isolated $tRNA^{Phe}$ from malignant human tissues was reported to contain more dihydrouridines (Kuchino and Borek, 1978). D is also present in urine samples and more significantly abundant in the urine of lymphoma patients (Jiang et al., 2016; Reimer et al., 1989). This is in line with the idea that tumor tissues undergo high turnover of tRNAs that can be quantified and used as a noninvasive biomarker for diagnosis and treatment of cancer (Borek et al., 1977; Dudley and Bond, 2014; Topp et al., 1993). The serum is also used to assess the metabolomic changes in various disease contexts; D is upregulated in the serum of patients with the major form of oral cancer in the world (oral squamous cell carcinoma) (Sridharan et al., 2017). At the molecular level, the human DUS2 protein (hDUS2) was shown to act as an interacting inhibitory factor of the interferon-induced protein kinase PKR – whose activity is enhanced in melanomas and colorectal cancers (Mittelstadt et al., 2008). Similarly, the anti-cancer ginsenoside compound was shown to repress the expression of *hDUS2* in human colorectal cancers cells (Luo et al., 2008).

To date, the most comprehensive study linking D with cancer was provided by Nakamura and colleagues (Kato et al., 2005). *hDUS2* showed a 3-fold overexpression in non-small cell lung carcinomas (NSCLCs) compared to healthy samples. The hDUS2 protein followed the same pattern, localized at endoplasmic reticulum and harbored a C-terminal double-stranded RNA binding motif (DSRM or dsRBD). The research cited above described an *in vivo* protein-protein interaction between hDUS2 and the glutamyl-prolyl tRNA synthetase EPRS and also identified *in vivo* an enhanced translational efficiency in the presence of hDUS2. Phenotypically, the suppression of the tumor cell growth was observed after siRNA-dependent *hDUS2* depletion and the NSCLCs patients with high levels of hDUS2 showed shorter survival periods. Their subsequent model was that overexpression of hDUS2 led to the hypermodification of tRNAs and consecutive increase of conformational flexibility. Because hDUS2 interacted with EPRS, they hypothesized that the tRNAs were more promptly charged in NSCLCs, which globally resulted in a more efficient translation. Strikingly, *hDUS1* was not differentially expressed in the same cell lines, raising the possibility of a hDUS2-specific regulation. However, this model is not fully supported by the data and requires more evidence; what is the biological relevance of the hDUS2-EPRS interaction? Isn't it an incidental interaction due to their common interaction with tRNAs? Because the hDUS2 overexpression does not necessarily mean that the D content is increased, are there more dihydrouridine residues in NSCLCs? *etc.*

d. Open questions

In the plant *A. thaliana*, D was not experimentally shown as a tRNA modification but was inferred to be synthesized due to the five putative *dus* genes encoded by the genome (Dus1, 2, 3 and DusA homologs, but never experimentally tested) (Kasprzak et al., 2012). Interestingly, these genes are lowly expressed in rosette leaves and apex tissues, unlike other RNA modifying enzyme genes (Chen et al., 2010).

The availability of numerous RNA-seq data allows an *in silico* assumption of the presence of RNA modifications that leaves detectable traces during library preparation (Findeiss et al., 2011; Ryvkin et al., 2013). Using this kind of pipeline, the plant transcriptome was assumed to carry D on lnc-, mi-, sno- and m- RNAs (Vandivier et al., 2015). Such predictions are poor evidence of the actual modification profile of a

transcriptome due to the weakness of the positive controls provided and the complexity to distinguish between more than 150 PTMs.

The reprogramming of yeast tRNA modifications in different conditions was thoroughly investigated by Dedon and others (Chan et al., 2015; Chan et al., 2010; Patil et al., 2012; Rose et al., 2016). D does not seem to be differentially synthesized on tRNAs when cells are exposed to various cellular stresses but is slightly more prevalent in the S-phase of the cell cycle. However, this effect is limited compared to the increase of mcm^5U (5-methoxycarbonylmethyluridine) (Patil et al., 2012).

In cardiomyocytes, the human DUS3 homolog was shown to interact with $polyA^+$ RNAs – along with 28 other RNA modification enzymes – raising the possibility of mRNAs hDUS3-specific modification in this specialized cell type (Liao et al., 2016). hDUS3 was also shown to be an inhibitor of the regenerative ability of the central nervous system (Buchser et al., 2012). Strikingly, the dihydrouridine was more abundantly detected during neural development in human embryonic stem cells (Basanta-Sanchez et al., 2016).

In *Trypanosoma brucei*, a carboxypropylated dihydrouridine was detected for the first time at the position 47 of $tRNA^{LysUUU}$ (**Fig. S4**). This new acp^3D RNA modification (3-(3-amino-3-carboxypropyl)-5,6-dihydrouridine) has an unknown structural function and its synthesis has still to be unraveled (Krog et al., 2011).

Given the biochemical peculiarities of dihydrouridine, it was included into *in vitro* prepared short interfering RNAs (siRNAs). The goal was to find out how to improve the pharmacokinetic features of siRNA duplexes for subsequent therapeutic utility (Wang et al., 2017). It turned out that the installation of a dihydrouridine at the 5'-end of the duplex increased its silencing activity (Sipa et al., 2007).

ii. Specificities in Bacteria

a. Bacterial D on tRNA and rRNA

Dihydrouridine is the most prevalent bacterial tRNA modification with pseudouridine (192 counts each). All canonical D residues (D_{16}, D_{17}, D_{20}, D_{20a}, D_{20b} and D_{47}) are found in Bacteria. Nevertheless, D_{47} is a scarce modification with a unique count described so far ($tRNA^{Met}$ of *B. subtilis*), although the position 47 in bacterial tRNAs is a U in almost 90% of the 134 known sequences. D_{20b} is also very uncommon and found on a cyanobacterial $tRNA^{Glu}$ (Machnicka et al., 2014).

A peculiarity of the bacterial dihydrouridine landscape is the presence of D on the 23S rRNA in Gram -negative (*E. coli*) and -positive (*M. hominis*, *C. sporogenes*) bacteria (Johnson and Horowitz, 1971; Kirpekar et al., 2018; Kowalak et al., 1995). These observations are of high interest because not a single dihydrouridine is reported on non-bacterial rRNAs. In *E. coli*, D turned out to be synthesized from U_{2449} and from U_{2500} in *C. sporogenes*, residues that are located in the highly conserved 23S rRNA central loop of domain V (Kirpekar et al., 2018; Kowalak et al., 1995; Popova and Williamson, 2014). Remarkably, this region is part of the peptidyltransferase center that is also one of the sites of interaction with antibiotics targeting ribosomal activity. However, D_{2449} is dispensable in *E. coli* and an unmodified pyrimidine at that precise

rRNA position is sufficient for viability (O'Connor et al., 2001). This suggests that the 23S-D_{2449} – instead of being an essential modification – could be involved in a fine-tuning regulation of translation in Bacteria.

The Gram-positive *C. acetobutylicum* is the only known organism to have a D on its 16S rRNA. The modification occurs on position 1211 or 1212 but its function is still unknown (Emmerechts et al., 2007).

b. Bacterial Dus enzymes

Taking advantage of the differential distribution of dihydrouridine across species, de Crécy-Lagard and colleagues implemented an *in silico* comparative genomic screen in order to find bacterial *dus* genes. Because no D was ever detected in *P. furiosus*, they assumed that the genome of this organism should not contain any *dus* gene, in contrast to other microorganisms such as *E. coli* or *S. cerevisiae*. By doing so, they found ortholog genes – or more precisely *clusters of orthologous groups* – that were absent in the D-free *P. furiosus* species but present in other D-containing species (Bishop et al., 2002). This approach led them to the identification of three *E. coli* Dus enzymes that are now called DusA, DusB and DusC. Importantly, they showed that the $tRNA^{Met}$-D_{20a} was DusA-specific and that the viable triple Δ*dusABC* mutant did not display any apparent growth defect and was devoid of tRNA dihydrouridines.

Similarly to eukaryotic Dus, bacterial Dus enzymes have non-redundant activities on tRNAs (**Fig. 7B**). *E. coli* DusB and DusC are mono-specific proteins that target U_{17} and U_{16}, respectively. The DusA substrates include U_{20} and U_{20a} (Bou-Nader et al., 2018).

c. Correlating D and growth temperature

The implementation of a quantitative detection of dihydrouridine led to the conclusion that psychrophilic bacteria have up to 70% more D on tRNAs than their mesophilic counterparts, which is in contradiction with the general observation that tRNAs of psychrophiles tend to be hypomodified (Dalluge et al., 1997). The psychrophilic organisms that grow between 0 and 20°C – with an optimum at 15°C – have the necessity to cope with low environmental temperatures, unlike the mesophiles bacteria that live above 20°C. It has been established that the lower the temperature, the higher the [C2'-*endo*/C3'-*endo*] ratio for dihydrouridine – this ratio representing the equilibrium constants as determined by NMR (Dalluge et al., 1996b; Deslauriers et al., 1971). In other words, low temperatures tend to stabilize the C2'-*endo* conformation of the dihydrouridine ribose moiety. The accumulation of D in psychrophilic prokaryotes could therefore constitute an evolutive adaptation to allow these organisms to maintain the conserved L-shaped conformation of tRNA, despite a growth at very low temperatures that otherwise could be detrimental for tRNA structure and function.

More generally, the set of modifications present on a tRNA depends on environmental cues, such as the temperature. In agreement with this idea, the *in vitro* synthesis of dihydrouridine in the hyperthermophilic bacterium *T. thermophilus* is possible on an unmodified $tRNA^{Phe}$ at 60°C but not at 80°C. At this latter temperature, the tRNA substrate needs to carry other modified nucleosides (Kusuba et al., 2015).

At the transcriptional level, the gene coding for the mesophilic *Clostridium botulinum* DusB homolog is downregulated during a heat shock stress at 45°C. In line with the above principles, the bacterium would require less D at high temperatures and would therefore decrease the expression of its cognate enzyme (Liang et al., 2013).

d. Open questions

Surprisingly, unexpected and putative Dus functions were observed, mainly in proteobacteria. *dusA* is upregulated in the β-proteobacterium *N. meningitidis* when in contact with epithelial cells and seems to play a role in bacterial adherence (Hey et al., 2013). The γ-proteobacterium *S. typhimurium* Y RNA – a ncRNA involved in RNA degradation – possesses a DusA-dependent D (Chen et al., 2014). *dusC* is downregulated in the γ-proteobacterium *E. coli* in an antibiotic-induced cell death (Kohanski et al., 2007). In other γ-proteobacteria, *dusB* is referred as a determinant for biofilm formation (Grasteau et al., 2011; Musken et al., 2010). The question is to know whether the modification itself is implicated in these biological processes. If not, it could suggest that bacterial Dus are multifaceted proteins that are not only RNA modifying enzymes.

iii. Specificities in Archaea

Archaebacteria are known to live in extreme conditions. Following the correlation established between D and temperature (see Introduction III.C.ii.c.), it is not surprising to observe that dihydrouridine is absent in most archaea. It is particularly evident for extreme thermophilic organisms such as *Pyrobacalum sp.* and *Pyrodictium sp.* whose optimal growth temperature is around 100°C. However, the *Thermococcus* species have dihydrouridylated tRNAs despite their optimal growth at 98°C (Edmonds et al., 1991). D was detected in other archaeal genera – although less temperature-extreme – such as *Methanococcus* (30°C), *Methanosarcina* (37°C), *Thermoplasma* (56°C) or *Methanobacterium* (65°C) (Best, 1978; Gupta and Woese, 1980; McCloskey, 1986).

An interesting case is found with *Methanococcoides burtonii* that has a thermal niche around 2°C and an *in vitro* optimal growth at 23°C. It was noticed that despite its tremendously low percentage of modified tRNA nucleotides – only 2%, which is one of the lowest in the living world – this psychrotolerant archaeon has, on average, more than one D residue per tRNA. Surprisingly, the D levels do not seem to vary with increased growth temperatures (from 4 to 23°C) (Noon et al., 2003). The fact that an archaeon living at low temperatures possesses D residues is another clue to consider this RNA modification as being involved in the flexibility of tRNAs when it is required to allow proper growth.

A biochemical study shed light on the temperature-dependent D-ring opening to 3-ureidopropionic acid (Fig. 5). The half-life of D at 37°C is about two hundred days and decreased to less than ten hours at 100°C (House and Miller, 1996). Extremophilic organisms either adopted a degradation/re-synthesis strategy to get rid of the 3-ureidopropionic acid-containing tRNAs or have acquired other molecular mechanisms that prevent the opening of the D-ring.

iv. A common feature: the Dus enzymatic mechanism

Dus are flavin-dependent enzymes that function similarly to the dihydroorotate and dihydropyrimidine dehydrogenases. Based on kinetics and structural data, an FMN- and NADP-dependent enzymatic mechanism has been proposed (Lombard and Hamdane, 2017; Rider et al., 2009; Yu et al., 2011); a hydride anion (H^-) is added to the nucleophilic C6 of uridine after the reduction of FMN to $FMNH^-$ by NADPH. The second hydride transfer to C5 likely occurs through the oxidation of a highly conserved Dus cysteine residue (Cys), that argues in favor of an evolutionary conserved mechanism (**Fig. 7C**) (Kasprzak et al., 2012; Savage et al., 2006).

v. Molecular evolution of Dus enzymes

An *in silico* high-throughput study has gathered 11,000 sequences of the Dus family for phylogenetic comparison (Kasprzak et al., 2012) (**Fig. 7D**). They identified eight sub-clusters, including the eukaryotic-exclusive Dus1, Dus2, Dus3 and Dus4 families; the bacterial-exclusive DusC family; the archaeal-exclusive Dus proteins; the DusA cluster found in bacteria and plants; and the DusB family that is mainly bacterial but also found in Archaea.

It is suggested that DusB is the oldest bacterial dihydrouridine synthase that evolved at a slow rate, unlike the rapidly evolving DusA. This latter is mainly found in proteobacteria and could therefore be the result of *dusB* duplication in the ancestor of this phylum. DusA orthologs found in other bacteria are likely products of horizontal gene transfer, whereas eukaryotic DusA-like proteins (in the fungus *Encephalitozoon cuniculi* or in the plant *Arabidopsis thaliana*) are probably the consequence of endosymbiosis.

In Eukarya, the mono-specific enzyme Dus3 (D_{47}) seems to be the most ancient Dus enzyme, followed by Dus2 (D_{20}). Subsequent duplications led to the production of enzymes with apparent relaxed specificity; Dus1 (D_{16} and $_{17}$) and Dus4 (D_{20a} and $_{b}$). It is noteworthy that DusB and Dus3 are both the oldest Dus proteins of their own domain and are closely related in terms of clusters projection, unlike DusA and Dus1-2-4 families that are highly different with one another.

Specific evolutive patterns were also observed. The *Bacillus subtilis* and *Clostridium acetobutylicum* firmicutes have two DusB paralogs encoded in their genome. As discussed above, these two bacteria are also characterized by unusual D modifications ($tRNA^{Met}$-D_{47} in *B. subtilis* and 16S-D_{1211} or $_{1212}$ in *C. acetobutylicum*). The presence of these noncanonical paralogs could explain the specific dihydrouridylation pattern of both species. Furthermore, DusC is missing in α-proteobacteria but data of tRNA sequences are still lacking and it is thus not known whether the DusC absence indeed results in non-modification of tRNAs-U_{16} (Fig. 7B). In the same bacterial class, *Rickettsia* genus seems to be deprived of DusA proteins. The reason of the *dusA* gene loss is unknown and the impact on the tRNA physiology should be investigated to bring answers about the evolutionary mysteries of Dus.

In conclusion, this bioinformatic-based approach revealed widespread distribution of dihydrouridine synthases throughout living organisms.

vi. Structural studies of Dus enzymes

To date, eight Dus enzymes structures have been published (**Fig. 7E**). The seminal crystallographic structure of an unknown FMN-binding protein in *T. maritima* revealed an oxidoreductase enzyme with two domains; an N-terminal TIM barrel and a C-terminal helical domain (Park et al., 2004). Later, this enzyme was referred as a dihydrouridine synthase. The *T. thermophilus* DusA crystal highlighted the same general structure (Yu et al., 2011). Moreover, the FMN cofactor (flavin mononucleotide) was captured in a positively charged groove at the center of the N-terminal domain corresponding to the catalytic site. DusA-tRNA^{Phe} complex revealed that DusA interacts with the DSL, ASL and TSL of tRNA (Fig. S3) and that the D-loop but not the D-stem is strongly distorted when DusA is bound. The third published bacterial Dus structure was DusC from *E. coli* that also displays a two-domain conformation with an N-terminal catalytic domain and a C-terminal RNA binding domain (Byrne et al., 2015; Chen et al., 2013). The structural similarities between *T. thermophilus* DusA and *E. coli* DusC led to the hypothesis that they share the same catalytic mechanism (see Introduction III.C.iv.). Remarkably, notable structural dissimilarities were discovered by comparing the bacterial DusA (targeting tRNA-U_{20} and $_{20a}$) and DusC (targeting tRNA-U_{16}). Both enzymes adopted the same general fold – while having different substrate specificities – but bound and recognized the tRNA in different orientations (**Fig. 7F**). The tRNA binding differed by a 160° rotation that resulted in the proper integration of the targeted uridine in the catalytic pocket. This trademark way of catalyzing a reaction is unique in RNA enzymology and is achieved through specific *binding signatures*. According to its target, each Dus enzyme has a cluster of amino acids – that is phylogenetically conserved in Bacteria – that defines the docking of tRNA to allow the reduction of a specific uridine (Byrne et al., 2015). The missing DusB structure was provided by Hamdane and colleagues (Bou-Nader et al., 2018). Even though the crystal was incomplete, it was concluded that *E. coli* DusB adopted the same overall structure with an N-terminal TIM barrel fold carrying the catalytic function and a C-terminal helical domain. The tRNA docking in DusB was similar to the one of DusC, which makes sense since DusB and C modify neighboring nucleotides (17 and 16). However, a major difference between DusB and C relied on the positioning of the nucleobase into the catalytic center. Reversed polar and nonpolar amino acids in the catalytic pocket of DusB led to the nucleobase rotation (180°C) that is targeted for reduction. In conclusion, diversification of bacterial Dus specificities was made possible through two astonishing strategies; nucleobase rotation or tRNA docking rotation (**Fig. 8**).

The only available eukaryotic Dus structure is the human DUS2. hDUS2 is particularly important because it is associated with pulmonary carcinogenesis and, unlike other Dus enzymes, harbors a dsRBD (see Introduction III.C.i.c.). This domain turns out to be conserved in animals (mammals, amphibia, flatworms, nematodes, insects) (Kasprzak et al., 2012). The sequence similarity is quite low between hDUS2 and bacterial Dus or even yeast Dus2, suggesting a potential novel strategy for the substrate recognition (Griffiths et al., 2012). The analogy between bacterial Dus and hDUS2 includes; (i) an N-terminal catalytic domain folded in a TIM barrel, (ii) an interaction of the catalytic domain with FMN, (iii) a high sequence conservation of the residues required in the active site – including the cysteine as an H^- donor and (iv) the presence of a helical domain (but not C-terminal). The dsRBD domain was shown to be necessary but not sufficient for D_{20} synthesis on yeast tRNA extracted from a Δ*dus2*

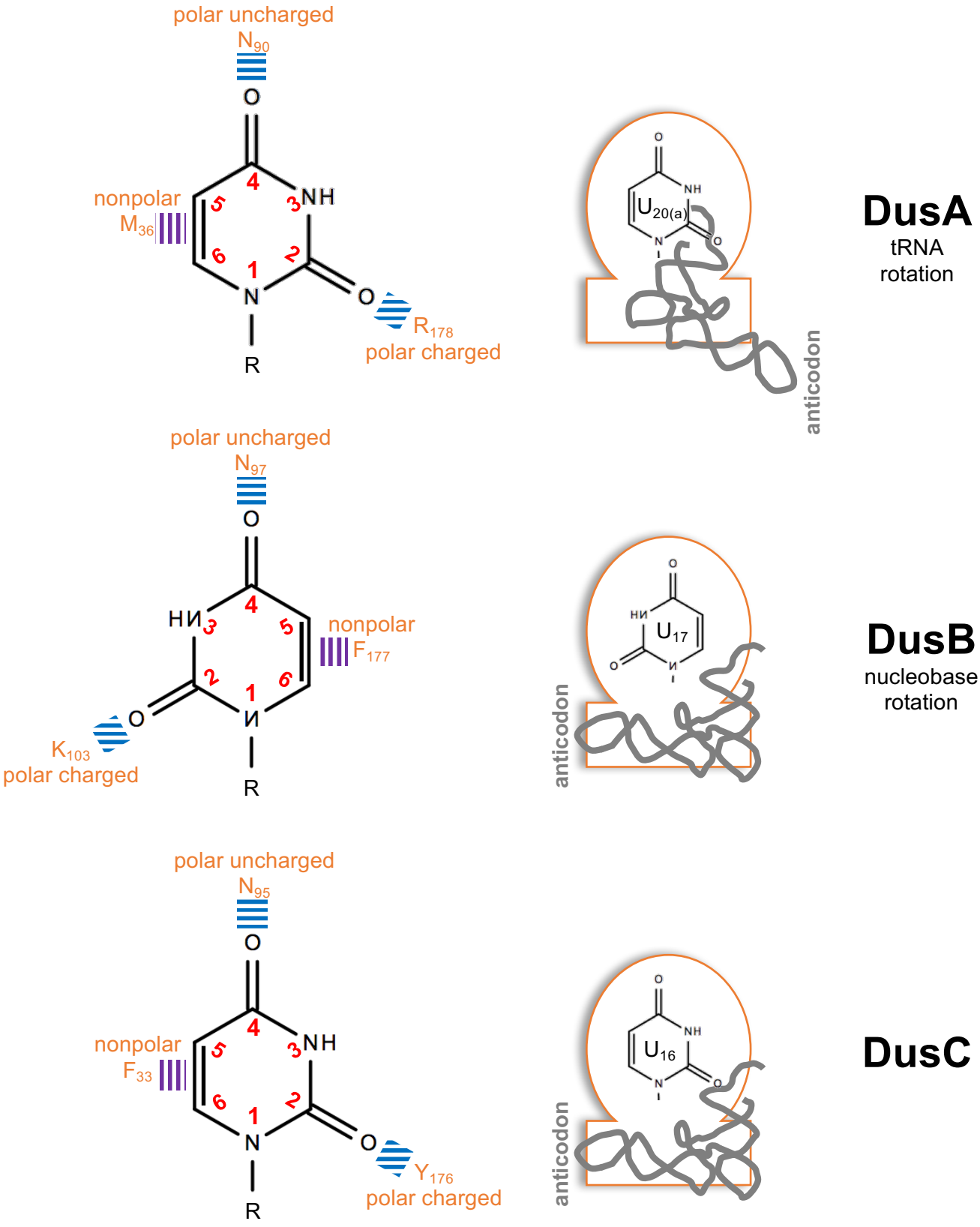


Figure 8 – Molecular strategies of bacterial Dus enzymes for substrate specificities. On the left, interaction of the U nucleobase (in the context of the RNA chain (R)) with Dus amino acids (orange: F for Phe, N for Asn, K for Lys and Y for Tyr) through ionic (blue lines) or hydrophobic (purple lines) interactions. In the DusB catalytic pocket, the rotation of the U nucleobase (180°) is observed in comparison to the DusA and C counterparts. **On the right**, schematic representations of the Dus enzymes (orange line) with the N-terminal nucleobase-containing catalytic domain (round) and the C-terminal RNA binding domain (rectangle). The tRNA (gray line) docking is similar in DusB and C but differs by a rotation (160°) in DusA. These strategies (tRNA or nucleobase rotations) allow the unvariable targeting of a specific uridine.

strain. Furthermore, this domain was suggested to serve as primary tRNA binding site, before the canonical helical domain (Bou-Nader et al., 2015). Using a dsRBD for tRNA recognition was never reported before, making hDUS2 the only known tRNA modifying enzyme to adopt this strategy.

vii. D and viruses?

Some viral genome-encoded tRNAs were shown to be dihydrouridylated on positions D_{16}, D_{17}, D_{20}, D_{20a} and/or D_{47} (Machnicka et al., 2014).

Tymoviruses are plant viruses (*e.g.* turnip yellow and eggplant mosaic viruses) and are known to adopt a noncanonical 3'-structure on their genomic RNAs, called *TLS* for tRNA-like structure (Rietveld et al., 1982). Despite the tRNA-like conformation with a D-loop/T-loop interaction, no dihydrouridylation was reported on these special structures (Becker et al., 1998; de Smit et al., 2002).

IV. THE DAWN OF A NEW ERA: EPITRANSCRIPTOMICS

Unlike the dihydrouridine-related works, the epitranscriptomic investigations have been recently and extensively reviewed (Harvey et al., 2017; Helm and Motorin, 2017; Kadumuri and Janga, 2018; Li et al., 2016b; Roundtree et al., 2017; Schaefer et al., 2017). The goal of this section is not to provide a comprehensive summary of the epitranscriptomic studies and breakthroughs, but mostly to introduce the main insights that are necessary to understand the concepts presented in this thesis.

A. Epitranscriptome?

[…] our understanding of an additional regulatory layer of biology that rests between DNA and proteins is still in its infancy; namely, the multitude of RNA modifications that together constitute the 'Epitranscriptome'.

(Saletore et al., 2012)

i. Definition

In 2012, the word *epitranscriptome* was used for the first time in a Cell resource (Meyer et al., 2012). The idea is that the *complete catalogue of RNAs present in a cell at a specific moment* (*transcriptome*) is far more than a set of polyribonucleotides made of A, U, G and C. There are, *on these building blocks* (*epi-*), a hidden layer of modifications that all together account for the *epitranscriptome.*

The idea of an *epitranscriptomic* world is conceptually similar to the epigenome (DNA modifications) and to the epiproteome (set of posttranslational modifications). But it is not revolutionary because; (i) RNA modifications have been known for 60 years (Davis and Allen, 1957) and (ii) the notion of a new type of biological regulation through the widespread distribution of RNA modifications had already been discussed (He, 2010). The scientific revolution lies on the possibility that the epitranscriptome; (i) is

dynamic and thus even more biologically relevant and (ii) is biotechnologically accessible, not only for one or two RNA species but at the *omics* scale.

An epitranscriptomic analysis is therefore defined as a transcriptome-wide mapping of an RNA modification of interest. In ten years (or less), an epitranscriptomic study will potentially be described as a transcriptome-wide mapping of several (all?) RNA modifications in a single experiment.

The Japanese expert Suzuki said “in the next ten years, I predict that the mapping accuracy of RNA modifications will improve, to the extent that it will become possible for researchers – irrespective of their discipline – to easily access databases containing up-to-date information on RNA modifications” ((Frye et al., 2016), p. 371).

ii. Epitranscriptomic studies

Understanding the functionality of RNA modifications is a hot topic in modern biology. To help unraveling these roles, the question *what for?* is not sufficient and needs to be supplemented with the question *where?*. Where are localized the RNA modifications on a transcriptome, and why?

To date, less than ten modifications have been studied at the epitranscriptomic level; m^6A, m^6Am, m^5C, hm^5C, ψ, m^1A and 2’-O-Me (reviewed in (Kadumuri and Janga, 2018)).

Besides the classic epitranscriptomic marks, there are other RNA modifications that are described at the transcriptomic scale. The internal inosine (I) modification was also identified in transcriptomic analyses either by comparing the DNA and RNA sequences (I is an A modification but is read as a G), or after chemical erasing by cyanoethylation (how a chemical treatment leads to PTM mapping is explained thereafter) (Li et al., 2009; Sakurai et al., 2010). The 5’-end modifications are other well-described and widely distributed modifications like the m^7Gppp cap on eukaryotic mRNAs. The first and second nucleotides that follow this conserved 5’-cap are often modified as well (sugar methylations and possible nucleobase methylation on the first nucleotide if it is an A) (Meyer and Jaffrey, 2014). In Bacteria, a novel 5’-structure was recently shown to have a widespread occurrence on small regulatory RNAs and some mRNAs. The covalently bound NAD cofactor (Fig. 3E) is dynamic and stabilizes targeted RNAs (Cahova et al., 2015).

B. Overview of the techniques

The major limitation of SBS [sequencing-by-synthesis] is that DNA polymerases are relatively blind to almost all DNA and RNA modifications […].

(Novoa et al., 2017)

i. Introduction

The direct sequencing of RNA molecules is still in its infancy and needs extensive improvements to become a reliable tool. The detection of RNA modifications is currently relying on the synthesis of a complementary DNA (cDNA) from any RNA

template that is – or not – decorated with chemical modifications. DNA is then sequenced using complementary fluorescent deoxyribonucleotides that are detected by an optical instrument. However, reverse transcription reactions (RNA → DNA) are usually insensitive to most RNA modifications. If not, discrimination between false positive and true positive hits is a painful task because the reverse transcription (RT) stops at multiple and non-relevant sites.

The transcriptome-wide mappings of PTMs are mostly achieved by implementation of antibody- or chemical-based techniques (subpoints ii. and iii.).

ii. Antibody-based epitranscriptomics

Monoclonal or polyclonal antibodies are obtained by the immunization of an animal and the targeted epitopes can be RNA modifications. Nevertheless, the production of reliable and highly specific antibodies against a modification of interest is burdensome, although efforts are made to expand the repertoire of antibodies (Feederle and Schepers, 2017; Novoa et al., 2017).

The antibody-based epitranscriptomics consists of the immunoprecipitation of RNAs that carry an RNA modification of interest. This protocol was first used in 2012 by two independent teams to map m^6A residues on mammalian transcriptomes (Dominissini et al., 2012; Meyer et al., 2012). RNA was extracted, randomly fragmented and then incubated with an anti-m^6A antibody. After immunoprecipitation (IP) of the methylated fragments, RNA was retro-transcribed into cDNA and the DNA library was prepared for next-generation sequencing (NGS) (**Fig. 9A**). The frequency of sequenced reads on and around a genomic position permitted the localization of m^6A residues. The subsequent bioinformatic analyses generated m^6A-peaks that represented the transcriptomic windows in which one or more methylated residues were found (**Fig. 9B**). This method was therefore not a single-nucleotide mapping because the resolution hovered around 200nt. However, the crosslinking-assisted sequencings enabled to narrow this quite large window to 23nt with PA-m^6A-seq and to the single-nucleotide resolution with miCLIP (Chen et al., 2015; Linder et al., 2015). PA-m^6A-seq stands for *photocrosslinking-assisted m^6A-sequencing* and includes a primary growth step with a uridine derivative (4-thiouridine), therefore allowing the crosslinking of the anti-m^6A antibody during the IP with fragmented RNA. This results in a decrease of the m^6A peak width along with an increase of the resolution around the putative methylated site (Chen et al., 2015). The miCLIP – *m^6A individual-nucleotide-resolution cross-linking and immunoprecipitation* – experimental pipeline also includes a crosslinking step but does not involve a uridine derivative. Here, the crosslinked anti-m^6A antibody is digested by a protease after the IP and leaves a trace that is detectable by sequencing. Because the trace takes place at a specific site (next to the modified residue), the mapping allows a single-nucleotide resolution (Linder et al., 2015).

The improvement of the antibody-based detection of m^6A also relies on a quantitative approach called m^6A-LAIC-seq for *m^6A-level and isoform-characterization sequencing*. For each sample, three sub-samples are sequenced; (i) the input RNA before IP, (ii) the supernatant harvested after IP and (iii) the immunoprecipitated fraction. Internal controls are added to all sub-samples and thus permit to determine

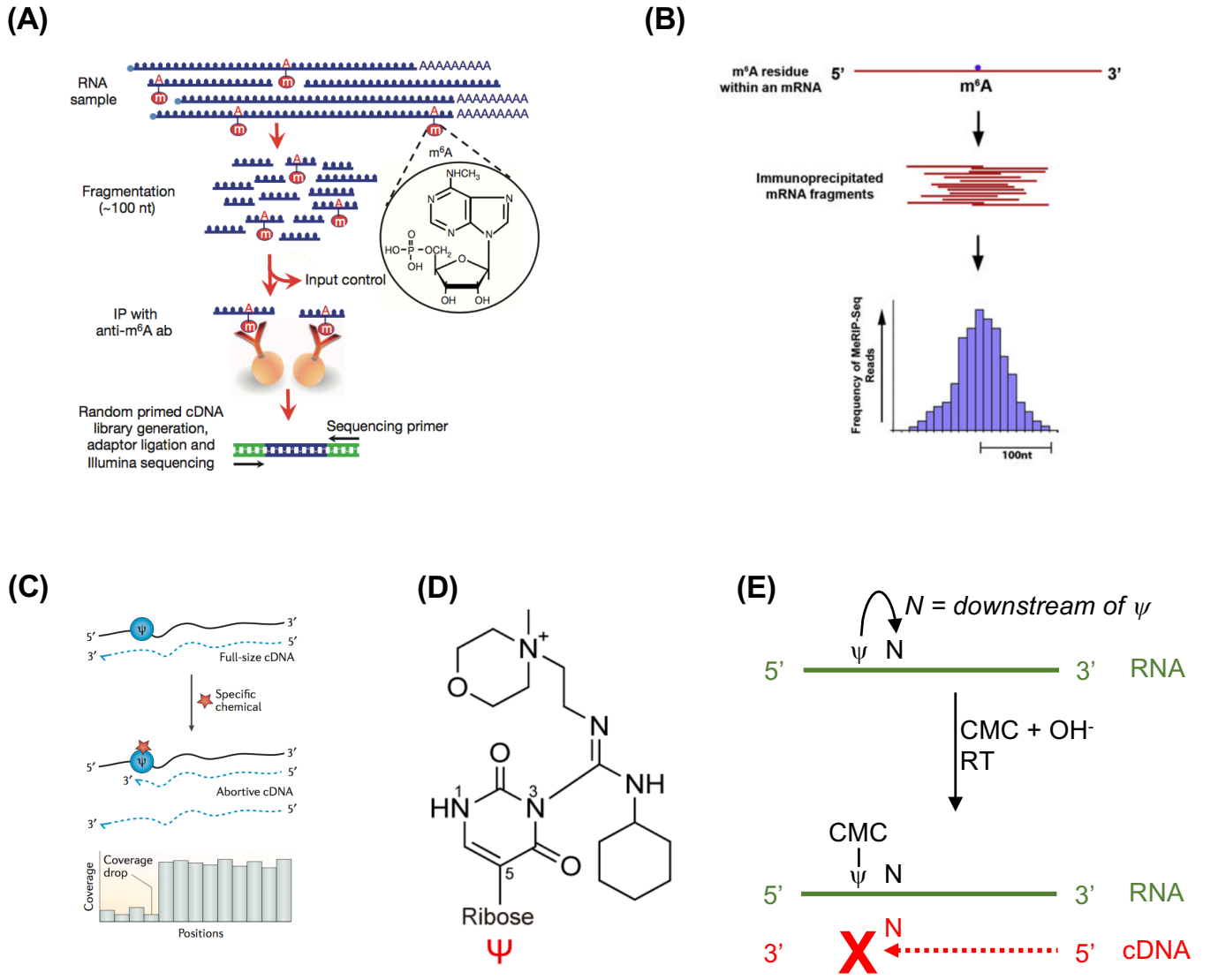


Figure 9 – Epitranscriptomics-related concepts. (A) Prinicple of immunoprecipiation (IP) of methylated RNA and subsequent DNA library preparation (m^6A-seq; Dominissini et al., 2012). **(B)** Enrichment of sequencing reads that correspond to m^6A-carrying RNAs and subsequent determination of a peak (purple) for the localization of the methylated adenosine (MeRIP-seq for m^6A-specific methylated RNA immunoprecipitation with next-generation sequencing; Meyer et al., 2012). **(C)** RT signature due to abortive reverse transcription after chemical treatment of a ψ-carrying RNA (Helm and Motorin, 2017). **(D)** ψ nucleobase is chemically modified by CMC after CMC and alkaline reactions (Song and Yi, 2017). **(E)** ψ marked with CMC results in RT termination one nucleotide downstream of ψ-CMC, as shown in Fig. 9C. The last retro-transcribed nucleotide is virtually not ψ, but N (A, U, G or C). The cDNA 3'-end sequence therefore indicates where RT was blocked. **(F)** Mispriming of the RT primer for cDNA production can lead to the misinterpreation of a modification of interest localization (Grozhik and Jaffrey, 2018).

precisely the stoichiometry of the modification for each mapped position (Molinie et al., 2016).

The immunoprecipitation-sequencing pipeline was applied to other RNA methylations; m^5C, hm^5C, m^1A and m^6Am (Delatte et al., 2016; Dominissini et al., 2016; Edelheit et al., 2013; Li et al., 2016a; Linder et al., 2015).

iii. Chemical-based epitranscriptomics

The chemical-based epitranscriptomic approach relies on the abortion of an RT reaction at a specific position, similarly to the concept of primer extension (Fig. S2), which was originally used to determine the 5'-end of a transcript. Reverse transcriptases (RTases) are RNA-dependent DNA polymerases that cannot produce DNA from any RNA template. The structural identity of the ribonucleic acid is indeed essential and impairments of the *normal* RNA architecture can lead to misincorporations into the DNA sequence, RTase pausing or even RT termination (**Fig. 9C**).

An RNA modification that strongly impairs the Watson-Crick interactions leaves *RT scars*, changes the cDNA sequence compared to the genomic information and is thus detected by sequencing (*e.g.* inosine) (Sakurai et al., 2010).

Most RNA modifications do not impair RT by themselves; a specific chemical reagent that *modifies the modification* is needed. For example, the two-step chemical treatment (CMC + OH^-) specifically labels ψ (**Fig. 9D**) (Song and Yi, 2017). The RTase does not process this heavily modified nucleobase, is blocked and thus leaves a truncated polymer. The cDNA 3'-nucleotide corresponds to the nucleotide downstream of the ψ residue (**Fig. 9E**). The epitranscriptomic mapping of ψ was performed by the implementation of the (CMC + OH^-) treatment, cDNA production and sequencing in the so-called ψ-seq, PSI-seq and Pseudo-seq experiments (Carlile et al., 2014; Lovejoy et al., 2014; Schwartz et al., 2014). The pipelines of these three concomitantly published studies were very similar and only differed in the preparation of the library for subsequent sequencing, in the controls and in the criteria used to call a pseudouridylated site.

Important note: the same nomenclature will be used throughout the manuscript and needs to be clearly indicated. When an RT termination is caused by the chemical labeling of an RNA modification at a position N, the RT-stop site is considered as the downstream nucleotide (N+1 or 3'-neighbor of the modification). The RNA sequence point-of-view is always taken into account. It also means that the modified residue is found one nucleotide upstream of the RT-stop site, because the cDNA point-of-view is never considered here. In other words, the cDNA 3'-end (or last retro-transcribed nucleotide) corresponds to the N+1 position of the RNA sequence and is described here as being downstream of the modified position. In conclusion, we consider that RT termination occurs one nucleotide downstream of the RNA modification.

RNA sequence: 5' - $N_{modification}$ - $N+1_{RT\text{-}stop\ site}$ - 3'
$N_{modification}$ = upstream of the RT-stop site
$N+1_{RT\text{-}stop\ site}$ = downstream of the modification

---

The use of a chemical reagent does not necessarily result in RT arrest; (i) after bisulfite treatment, cytidines are converted to uridines, whereas the m^5C residues remain unaffected. Deep sequencing reveals these differences for m^5C mapping

(Schaefer et al., 2009); (ii) after mild alkaline treatment, the RNA chain is partially cleaved but less significantly at 3'-phosphodiester bonds of 2'-O-methylated nucleotides because they are more resistant to a nucleophilic attack. A consequence of this *protection* is that the position +1 to the methylated residue is virtually never found at the 5'-end of fragmented RNA. The nucleotides that are under-represented at the 5'-ends are then detected by a lack of signal in deep sequencing. Differential RNA cleavage profiles are therefore analyzed in the epitranscriptomic approach called RiboMethSeq (Birkedal et al., 2015; Marchand et al., 2016).

Yet to be applied specific chemical reactions followed by high-throughput sequencing exist and the implementation of one of them is the purpose of this work (Behm-Ansmant et al., 2011; Helm and Motorin, 2017).

C. Concerns about the epitranscriptomic studies

[…] mapping results have been fundamentally challenged, raising doubts about the integrity of the field.

(Grozhik and Jaffrey, 2018)

The reliability of the epitranscriptomic investigations has recently been questioned in several issues (Grozhik and Jaffrey, 2018; Helm and Motorin, 2017; Schwartz and Motorin, 2017). First, the seminal antibody-based mappings of m^6A used an antibody that was not fully specific for the modification of interest, resulting in detection of false-positive hits (Linder et al., 2015). Then, three examples have puzzled the scientific community about the main conclusions on transcriptome-wide studies.

i. m^1A: from 7,000 to 15 sites

In 2016, 7,154 transcriptomic positions were assigned as m^1A RNA modifications in mammals and m^1A was shown to be a prevalent dynamic internal mRNA modification in the fission yeast (Dominissini et al., 2016; Li et al., 2016a). One year later, Schwartz and colleagues drastically improved the protocol, reported only fifteen m^1A-sites on the human transcriptome and revealed that the methylation of the mitochondrial *ND5* mRNA was developmentally regulated (Safra et al., 2017). The marked discrepancy (up to 500-fold) between the published epitranscriptomic profiles relies on the fact that the last study utilized not only one RT signature (truncated cDNA – Li et al., 2016a – or misincorporation at m^1A sites – Dominissini et al., 2016) but combine both experimental procedures to reliably detect m^1A residues.

ii. 2'-O-Me: the mispriming artifact

Whereas RiboMethSeq relies on the lack of signal at 5'-ends of RNAs after alkaline treatment (5'-N_m-N_{+1}-... and not 5'-N_{+1}-…) to map ribose methylation (see Introduction IV.B.iii), the detection of a positive signal has also been used in Nm-seq (or 2'-O-methylation-sequencing) (Dai et al., 2017). In this case, the periodate oxidation is used to specifically protect 2'-O-methylated nucleotides from digestion in the 3'-to-5' direction. As a consequence, the 3'-end nucleotide is more likely to carry a 2'-O-Me. The precise determination of the RNA 3'-termini is thus essential for proper sugar methylation mapping. To do so; (i) an adapter is ligated to the 3'-end of RNA

fragments, (ii) the RT primer is annealed to the adapter and (iii) the first retro-transcribed nucleotide corresponds to the 3'-end nucleotide of the RNA fragment which, in this case, is expected to be 2'-O-methylated. Mispriming of the RT primer would thus be dramatic for the correct mapping of modified riboses (**Fig. 9F**). This mispriming artifact is likely to have occurred in the work cited above according to the American expert Jaffrey who highlighted the fact that the consensus sequence of 2'-O-Me was highly similar to the sequence of the 3'-adapter (Grozhik and Jaffrey, 2018).

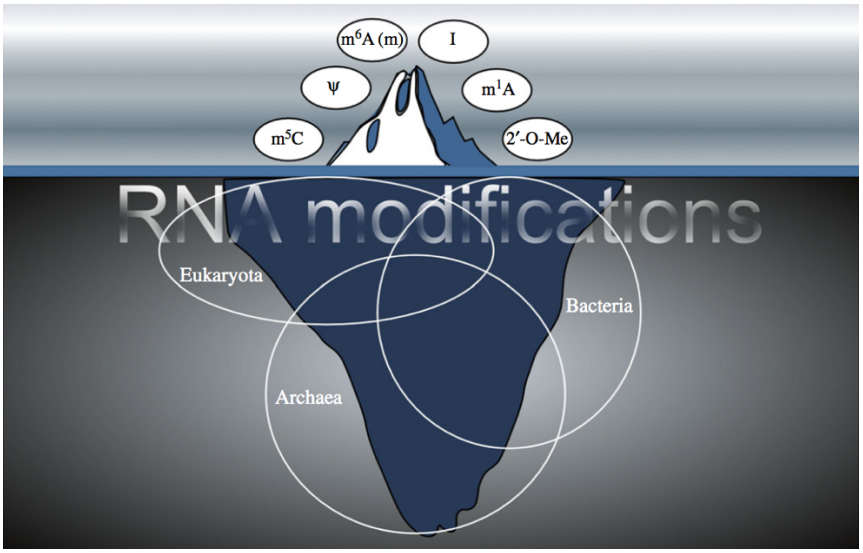
iii. m^{5}C: challenging the chemical reaction effectiveness

m^{5}C detection is based on the fact that it is insensitive to bisulfite reaction (see Introduction IV.B.iii). Although it is a true statement, it is not sufficient to overcome two major technical boundaries that created concerns about the m^{5}C epitranscriptomic reports; (i) other RNA modifications may be resistant to this *specific* treatment (*e.g.* N^4-methylcytidine) (Squires et al., 2012) and (ii) all cytidines are not converted to U after the bisulfite reaction. Even if only 0.2% of C are not converted, it can represent up to 20,000 incorrect modification callings in a transcriptome containing 10 millions of cytidines (Helm and Motorin, 2017).

iv. Short conclusion

The few epitranscriptomic-associated pitfalls presented in this section highlight the necessity to find strategies, controls and validations in order to enhance the true positive called sites along with the reduction of sequencing noise.

OBJECTIVES



(Schaefer et al., 2017)

D is a universal RNA modification whose abundance is not equally distributed across domains of life. Such a conserved modification is not common in the world of posttranscriptional modifications and is therefore of high interest. Furthermore, the gene coding for the *ancestral* dihydrouridine synthase was duplicated – up to three times in Eukarya – throughout the story life, suggesting a beneficial role for the global fitness of living organisms.

The apparent complete lack of dihydrouridine was achieved by genetic manipulation of two unicellular model organisms; the bacterium *Escherichia coli* and the fungus *Saccharomyces cerevisiae*. The loss of this RNA modification did not result in any obvious phenotype, as it is often the case when playing with RNA modification enzymes. Nevertheless, structural biologists provided insightful evidence indicating that the dihydrouridine strongly impacts the molecular thermodynamic equilibrium and brings flexibility to ribonucleic acids by promoting uncommon nucleotide conformations. In agreement with these observations, microbiologists detected a high prevalence of dihydrouridine in psychrophilic prokaryotes, but not in thermophilic archaea, providing a first clue to unravel the physiological role of this RNA modification.

However, a set of intriguing and still elusive facts about dihydrouridine and its cognate enzymes shed light to the importance of investigating the biological relevance of this modification; overexpression of Dus2 in pulmonary carcinogenesis, potential role of DusA in bacterial adherence, interaction of Dus3 with polyA-RNAs in cardiomyocytes, subtle role of a dihydrouridylated position on bacterial rRNA, *etc.*

The cutting-edge technologies such as the next-generation sequencing opened the door to the birth of an exciting field in molecular biology; the epitranscriptomics. By the implementation of epitranscriptomic layouts, the profiles of transcriptomic marks became accessible. Despite this novel available biotechnology, many technical challenges have yet to be overcome; this is reflected by the fact that less than 5% of

the currently known RNA modifications have been studied at the epitranscriptomic level.

We aim at setting up an easy-to-use and reliable technology in order to study the transcriptomic distribution of dihydrouridines. Our goal is to provide a powerful technique to bring new insights about the physiological relevance of D residues. Unbiased approach – such as an epitranscriptomic profiling – is essential to unravel the putative biological function of dihydrouridine beyond transfer RNAs. Our purpose is also to apply this approach to different model organisms to widen our conclusions with a phylogenetic point-of-view.

Our interest on dihydrouridine results – as always in science – from a completely different research story. By studying the role of a long noncoding RNA in fission yeast (Fauquenoy et al., 2018), we detected Dus2 as an interacting partner of this RNA in a pull-down assay (unpublished data). We therefore hypothesized that a dihydrouridine-dependent regulation acted on this lncRNA. But to decipher this potential mechanism, we first needed an answer to the question *where does the dihydrouridylation occur?*

In summary, the epitranscriptomics is an on-hand resource and should therefore be implemented to the most prevalent RNA modifications, such as the dihydrouridine.

RESULTS

Dihydrouridine is fluorescently-labeled with rhodamine.

A D-specific chemical reaction was used to detect the presence of D residues on RNA. In a two-step protocol, total RNA was first incubated with sodium borohydride ($NaBH_4$) and then covalently bound by the rhodamine stable fluorophore (Rho) (Fig. 6). Uridines are $NaBH_4$-insensitive and were thus not bound by the fluorophore. We used rhodamine labeling (Rho labeling) for the specific detection of D in two experimental layouts; (i) R+ ($NaBH_4$ + Rho) and (ii) mock or R- (KOH + Rho). Dot blot assays revealed that RNA-Rho wass readily detectable at 520nm and that Rho did not aspecifically bind RNA in mock-treated samples. Moreover, we generated a quadruple mutant for the four putative dihydrouridine synthases (Dus) in fission yeast (Δ*dus1-2-3-4* or Δ*4*). R+ labeling on total RNA from Δ*4* showed a decreased signal intensity compared to WT RNA (**Fig. 10A**). The remaining signal in Δ*4* demonstrated that D was not the only RNA modification to be labeled by Rho, as already shown by others (see Discussion and Perspectives). This observation was confirmed by electrophoresis of R+ and mock-treated RNAs and subsequent fluorescent detection (**Fig. 10B**). rRNAs were undeniably labeled although D was never reported as a eukaryotic rRNA modification, indicated the targeting of other RNA modifications by R+ treatment. Moreover, the rRNA was also labeled in a Δ*4* R+ background, suggesting a D-independent signal. Nevertheless, no fluorescence was observed in R- samples, demonstrating that Rho binding and detection are not due to secondary structures. Strikingly, a sharp decrease of Rho signal was observed on Δ*4* R+ tRNAs, as expected. Finally, the methylene blue staining used as a loading control revealed a partial degradation of both R+ and R- RNAs, as confirmed by microchip analysis (**Fig. S5**).

These data outline a D-dependent but not D-exclusive Rho labeling of RNA. Working with a quadruple *dus* mutant is thus required to discriminate between D-dependent and D-independent signals in WT R+ total RNA.

Dihydrouridine is mapped at the single-nucleotide resolution.

A primer extension assay was performed on $tRNA_i^{Met}$ to test Rho labeling on the known D_{16} position, putatively Dus1-dependent. The reverse transcription reaction (RT) on WT R+ RNA generated a 30nt-long radioactively labeled polynucleotide corresponding to the occurrence of RT termination one nucleotide downstream of D_{16}. WT R- and more importantly Δ*dus1* R+ and R- RNAs were devoid of the 30nt-polynucleotide (**Fig. 10C**). The *in vitro* dihydrouridylation on Δ*dus1* RNA with GST-purified Dus1 (but not Dus2 and Dus4, that were shown to be catalytically active in other assays) restored the RT termination downstream of D_{16}. Combining fission yeast-Δ*dus1* RNA and budding yeast-Dus1 protein also resulted in the restoration of RT arrest (**Fig. 10D**). The primer extension assay performed on a putative Dus2 tRNA target led to the same conclusions (**Fig. 10E**).

An *in vitro* transcribed synthetic RNA was then used to confirm RT termination one nucleotide downstream of a D residue (**Fig. 10F**). The DNA sequence possessed a unique T and the T7-dependent *in vitro* transcription was performed with uridine or commercially available dihydrouridine. The resulting U-containing or D-containing

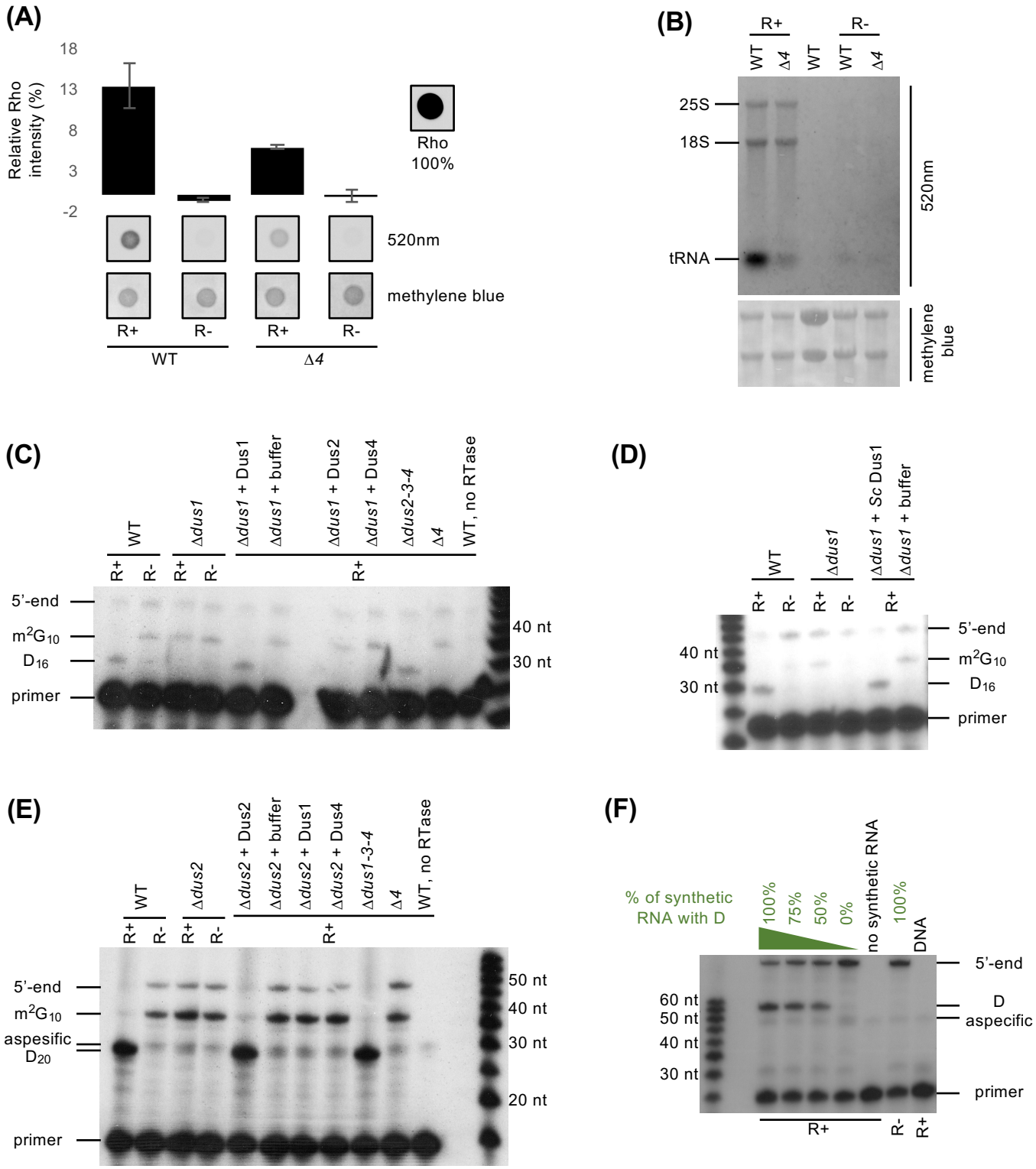


Figure 10 – Single-nucleotide detection of dihydrouridine (D) by rhodamine labeling. (A) Comparative Rho (rhodamine) detection at 520nm on WT (wild type) and Δ*4* (Δ*dus1-2-3-4*) total RNAs in R+ and mock-treated (R-) samples. The intensity of a Rho drop is scaled at 100%, a drop of water is considered as the background, methylene blue staining serves as loading control, *n* = 3, error bars represent SEM. **(B)** Fluorescent Northern blotting of WT and Δ*4* total RNAs in R+ and mock-treated samples. Untreated total RNA from WT strain is shown in lane 3. Representative result from a biological triplicate, methylene blue staining serves as loading control. **(C-D-E)** Primer extension assays performed on total RNAs from WT, single *dus* (Δ*dus1*, Δ*dus2*), triple *dus* (Δ*dus1-3-4*, Δ*dus2-3-4*) or quadruple *dus* (Δ*4*) mutant strains, R+ or R- treated. Before the chemical reaction, *in vitro* dihydrouridylations are performed with *S. pombe* GST-purified proteins (+ Dus1, 2 or 4), *S. cerevisiae* GST-purified Dus1 (+ *Sc* Dus1) or no protein (+ buffer) on *S. pombe* total RNA from different genetic backgrounds. The ladder consists of ^{32}P-labeled DNA oligonucleotides ranging from 15 to 60 nt. (C-D) primer specific for $tRNA_i^{Met}$, (E) primer specific for $tRNA^{AspGUC}$. **(F)** Primer extension assay performed on an *in vitro* transcribed RNA mixed with *S. pombe* total RNA. In green, decreasing amount of dihydrouridylated and increasing amount of uridylated synthetic RNAs are added before R+ or R- chemical reactions. DNA is the PCR product used for *in vitro* transcription.

RNAs were mixed in different proportions, R+ labeled and retro-transcribed. The presence of D was absolutely essential to generate RT termination one nucleotide downstream of the modified residue in R+ conditions. The synthesis of a synthetic RNA with three consecutive D generated a strong RT arrest downstream of the third dihydrouridylated site in the R+ condition (**Fig. S6**). This effect was dissipated for the following modified residues. Interestingly, a sharp RT termination was observed after the three D when they were unlabeled, strengthening the idea that the D modification itself can disturb the reverse transcriptase, and even more in a D-rich context.

Altogether, these data show that R+ labeling allows the single-nucleotide mapping of D. Moreover, Dus enzymes specificities on tRNAs (Fig. 7A) seem to be conserved in fission yeast.

Transcriptome-wide mapping of D by dihydrouridine-sequencing.

We developed D-seq (dihydrouridine-sequencing) to map D across the transcriptome. The protocol was performed as follows; (i) RNA extraction from WT and *Δ4* strains, (ii) R+ and mock-treatments, (iii) ribodepletion, (iv) cDNA synthesis and library preparation, (v) strand-specific deep-sequencing, (vi) data analysis by implementation of a multifactorial analysis (**Fig. 11A**).

The classic library preparation for RNA sequencing does not fit with the purpose of D-seq due to the lack of RT termination location. We therefore implemented an RNA preparation that permitted the determination of cDNA 3'-ends (**Fig. 11B**).

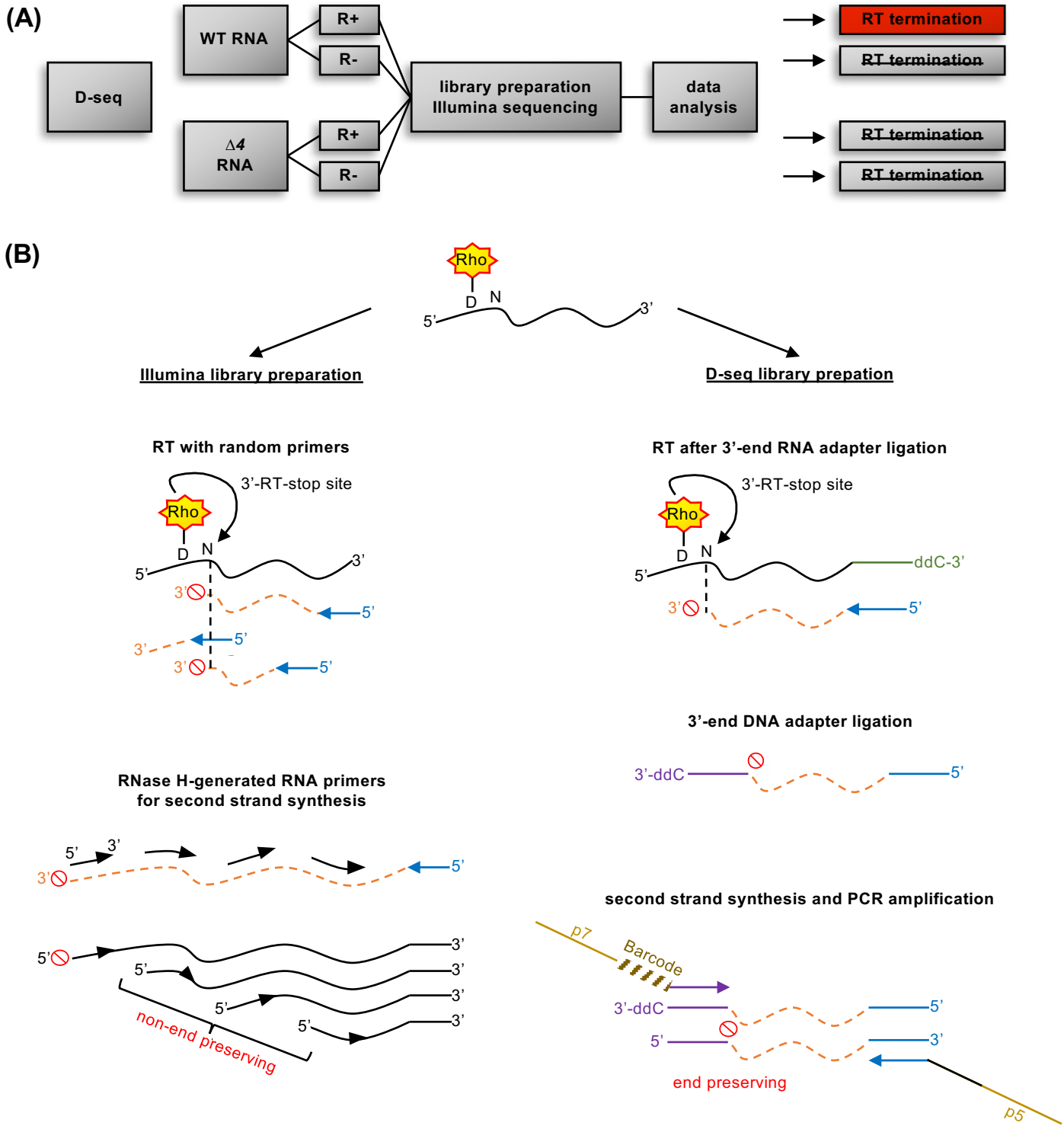


Figure 11 – Transcriptome-wide mapping of D by dihydrouridine-sequencing (D-seq). (A) D-seq general pipeline starting with total RNA (in this case originating from WT and Δ*4* yeast strains) and leading to the detection of RT termination events (red; WT R+ or test condition) or not (gray; WT R-, Δ*4* R+, Δ*4* R- or control conditions). **(B)** Comparison of library preparation strategies from a hypothetical RNA molecule (black line) carrying a Rho-labeled D residue upstream of a nucleotide N (A, U, G or C). In both cases, the blue arrows indicate the DNA primers used for the production of cDNA (orange dashed lines) by reverse transcription (RT). The vertical black dashed lines represent RT-stop sites, positions at which the RTase is blocked. The RT premature termination signatures on cDNA 3'-ends are indicated with red strikethrough circles. **On the left** (classic Illumina preparation), the synthesis of the second DNA strand relies on the enzymatic digestion of the RNA template to form oligoribonucleotides that will act as primers for the cDNA-dependent DNA synthesis. As indicated in the cartoon, this step leads to the loss of the cDNA 3'-end localization information. **On the right** (D-seq preparation), a 3'-end blocked RNA adapter (green line, ddC stands for dideoxycytidine) is ligated to the RNA of interest and is used for hybridization of a non-random RT primer. After the cDNA production, a second ligation is performed by addition of a 3'-end blocked DNA adapter (purple line) allowing the *imprisonment* of the cDNA 3'-end for subsequent determination of its precise positioning (corresponding to the position where the RTase was blocked). The second DNA strand synthesis is done along with PCR amplification. p5 and p7 indicate sequences that anneal to the Illumina flow cell.

We then based our analysis on the ψ-seq two-metric system and improved it to map positions at which the RTase signal dropped off in a D-dependent manner during cDNA synthesis. After the mapping of the reads on the genome, we compared the coverage of each position in the four sequenced conditions, that is to say WT R+, WT R-, Δ*4* R+ and Δ*4* R-. The test condition will always refer to the WT R+ condition whereas the control conditions encompass the three WT R-, Δ*4* R+ and Δ*4* R- conditions.

Our hypothesis was based on the fact that a portion of RT-stop sites (corresponding to the cDNA 3'-end nucleotides) should be downstream nucleotides of a D position in the test condition but not in the control conditions. Our goal was therefore to find the RT-stop sites that are specific to the test condition (meaning that they are labeling- and *dus*-dependent) to extrapolate about the positioning of an upstream D position.

A multifactorial analysis was performed to determine a nucleotide as an RT-stop site (**Fig. 12**); (i) a pre-filtering analysis with the calculation of an average D-ratio that was defined as the proportion of reads attesting RT termination out of all reads overlapping the position of interest. A minimum standard deviation of this ratio is required to ensure that all values were not similar; (ii) a statistical analysis to assess the interaction between test and control conditions and (iii) a post-filtering calculation of a D-fold change (D-fc) that was the logarithmic transformation of D-ratio by comparing the test (where an RT termination was expected) and the control (where an RT termination was not expected) conditions. This three-step analysis resulted in a high confidence determination of D-dependent RT-stop sites for which the upstream nucleotide (with the RNA sequence point-of-view) was considered as a dihydrouridylated position (D-site). The D-site was thus defined as a putative dihydrouridine. In summary, our bioinformatic pipeline included; (i) the positioning of labeling- and *dus*-dependent RT-stop sites, reflecting the single-nucleotide resolution of RT termination events in the WT R+ condition, and (ii) the mapping of D-sites that were located upstream of the RT-stop site.

To test the ability of Rho labeling to cause RT termination one nucleotide downstream of a labeled D in a deep sequencing context, a synthetic spike-in RNA was added to the samples (the same as used in Fig. 10F). The single D of the *in vitro* transcript generated an RT-stop site one base downstream of the modified base in an R+ sample (D-ratio 17.5%), but not in a mock-treated RNA (D-ratio 1.0%) (**Fig. 13A** and **B**). Importantly, an un-dihydrouridylated spike-in (carrying a single U) displayed a D-ratio of only 0.1% at the potential RT-stop site in both labeling conditions (**Fig. 13C** and **D**). Mixed populations of U and D spike-ins in R+ samples revealed a high correlation (R^2 = 0.94) between the RT-stop site D-ratios and known stoichiometries of D (from 0 to 100%) (**Fig. 13E**), but not in R- treated conditions (**Fig. S7**). It is noteworthy that a fully dihydrouridylated transcript (100% of D spike-in) induced an RT termination for only one out of five RT reactions (D-ratio ~ 20%), giving an important insight about the stoichiometry of coupling Rho labeling and sequencing. It indicated the limitation of one or several steps of the protocol, such as partial $NaBH_4$ reduction of D, partial Rho incorporation and/or ability of the RTase to overcome a *rhodaminized* residue. However, a D-ratio of 20% was more than enough to discriminate actual RT termination from the general background. Importantly, misincorporation by the RTase

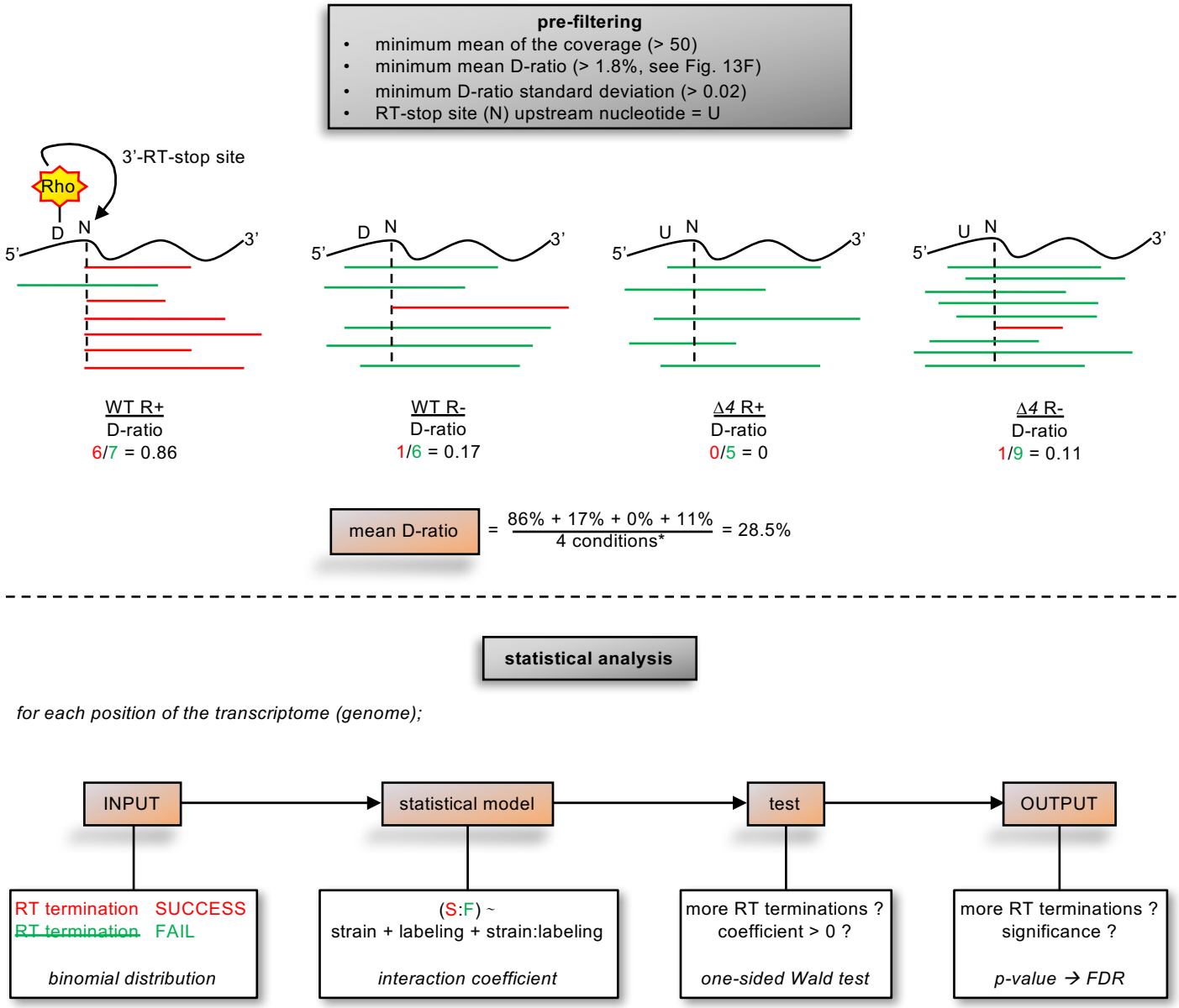


Figure 12 – D-seq analytical pipeline. The D-seq analysis is divided into three main sub-categories; (i) a **pre-filtering** analysis consists of an independent determination of reads supporting RT termination (red lines) out of all reads (red + green lines) overlapping the position of interest (N). This value is called the D-ratio and is calculated for all conditions. A mean D-ratio is obtained for each position of the transcriptome. Asterik notifies that there are basically eight (not four) D-ratios that are included in the mean D-ratio (4 conditions x 2 replicates). The pre-filtering also includes other features that are detailed in the shaded frame, (ii) the **statistical analysis** aims at answering to the question of RT termination occurrence significativity in the test condition (WT R+) compared to the control conditions (WT R-, Δ*4* R+ and Δ*4* R-). To test whether the potential increase of RT stop events in the test condition is strain- and labeling-dependent, an interaction coefficient is calculated. Because it is expected to have more RT arrests in WT R+ (> 0), a one-sided Wald test is thus performed and the resulting *p*-value is corrected to FDR (false discovery rate), (iii) the **post-filtering** analysis compares test and control conditions through a logarithmic transformation of the D-ratios. A minimum D-fc (fold change > 2) ensures that the mean D-ratio of the WT R+ is at least four times bigger than the mean of all control D-ratios.

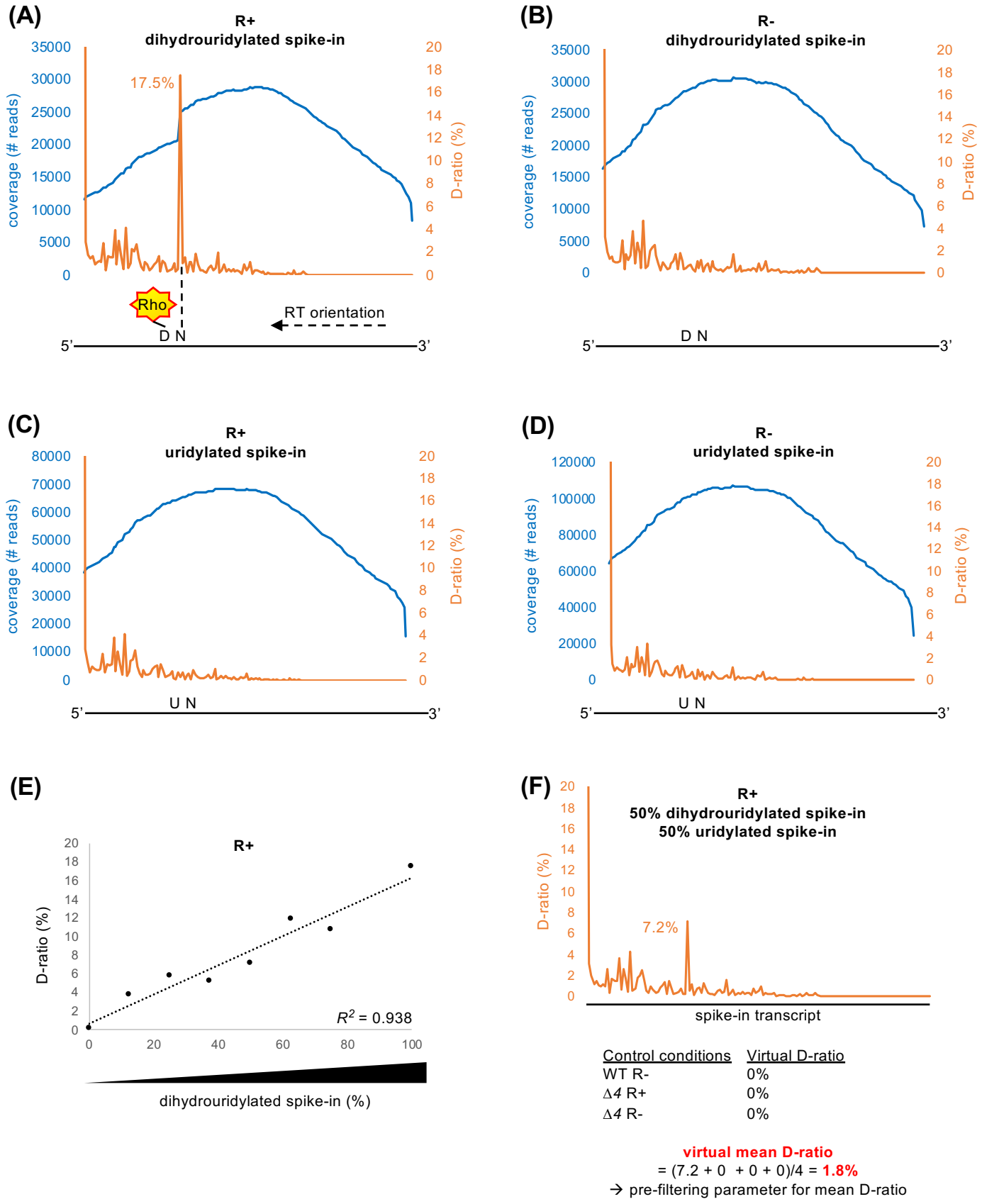


Figure 13 – Evaluation of the relative dihydrouridylation stoichiometry . (A-B-C-D) Evolution of the D-ratio (%, orange) along the synthetic RNA. Spike-in was mixed with total RNA before R labeling and deep sequencing. Coverage is shown in blue. (A) The D spike-in in the R+ reaction generated a high D-ratio one nucleotide downstream (position 45) of the unique D present in the sequence (position 44). (B) The D spike-in in a mock reaction did not result in RT termination neither at the position 45 nor at the position 44. (C-D) The U spike-in was free of RT arrest around the unique U of the sequence in both chemical reactions. **(E)** The D-ratio of the spike-in position 45 gradually increased along with D:U content in R+ condition. 100% means that all spike-in RNAs added to the sample carried a single D whereas 0% corresponds to a U spike-in. Correlation coefficient (R^2) close to 1 indicates a correlation between the parameters displayed in the chart. **(F)** Determination of a mean D-ratio threshold for the D-seq analysis (Fig. 12). When half of the spike-in RNAs are dihydrouridylated, RT termination at position 45 occurred for 7.2% of the reads in the R+ condition. This percentage is expected to be 0 in control conditions, resulting in a virtual D-ratio of 0.018 for a position modified with a 50:50 ratio.

(of a C, G or T) at the D position was not observed within the 80% of cDNAs that did not stop downstream of the dihydrouridine (data not shown).

Because a 50:50 proportion of D:U spike-in in R+ condition led to a D-ratio of 7.2% one base downstream of the D/U position, we hypothesized that a transcriptomic position that is dihydrouridylated in 50% of the molecules should have a mean D-ratio of 1.8% (**Fig. 13F**). We chose this value as a requirement to consider a position as a D in the pre-filtering analysis (Fig. 12).

Ability of D-seq to map dihydrouridylated residues is effective at a single-nucleotide resolution, as shown with the compelling spike-in control.

D-seq is a reliable tool to detect D positions on yeast tRNAs.

Taking advantage of the well-described modification profiles of tRNAs, we detected 228 D-sites on 141 yeast cytoplasmic (87%) and mitochondrial (13%) tRNAs. Two randomly chosen positions were confirmed by primer extension assays (**Fig. 14A**). Erasing the condition '*RT-stop site upstream nucleotide = U'* in the pre-filtering analysis (Fig. 12) resulted in the detection of 264 putative D-sites (instead of 228) where 86% were U and only 14% were G or A (**Fig. S8**). However, among these 36 new candidates (for which the D-site ≠ U), 94.4% had a U at the RT-stop site itself, strongly suggesting that the RTase terminated at the dihydrouridylated position rather than one nucleotide downstream. In order to avoid complexification of the analysis and potential false-positive candidates, we did not take the '*RT-stop site = D'* scenario into account.

In *S. pombe*, nine D positions have been experimentally reported on tRNAs (Keith et al., 1993; McCutchan et al., 1978; Vogeli, 1979; Wong et al., 1979; Zallot et al., 2014). The D-seq analysis detected six of them (**Fig. 14B**). Strikingly, 98.7% of all detected D were localized on the D-loop, as expected (**Fig. 14C**). The two most prevalent cytoplasmic tRNA positions highlighted by our method were D_{16} and D_{20}, which is in total agreement with the relative D occurrence in cytoplasmic tRNAs of another yeast (**Fig. S9**) (Xing et al., 2004). It is not surprising that we did not detect any D_{20b} or D_{47} on cytoplasmic tRNAs because they are rare modifications. Moreover, D_{47} is localized near the tRNA 3'-end and a cDNA product that would be terminated at D_{47} would be 20 to 50nt-long, which is a too short fragment to be considered in D-seq due to a size selection during the library preparation.

Using a 32nt-long window, no consensus sequence was clearly observed around the putative D residues on cytoplasmic tRNAs. As already reported for cytoplasmic tRNAs, a GG dinucleotide was enriched upstream of the modification (**Fig. 14D**). We reasoned that because Dus enzymes have high substrate specificities, the potential consensus sequences should vary with the different targets. No substantial differences were highlighted between Dus1 (D_{16} and $_{17}$), Dus2 (D_{20}) and Dus4 (D_{20a}) targets, arguing that the installation of D in tRNAs is structure- rather than sequence-dependent (**Fig. S10**).

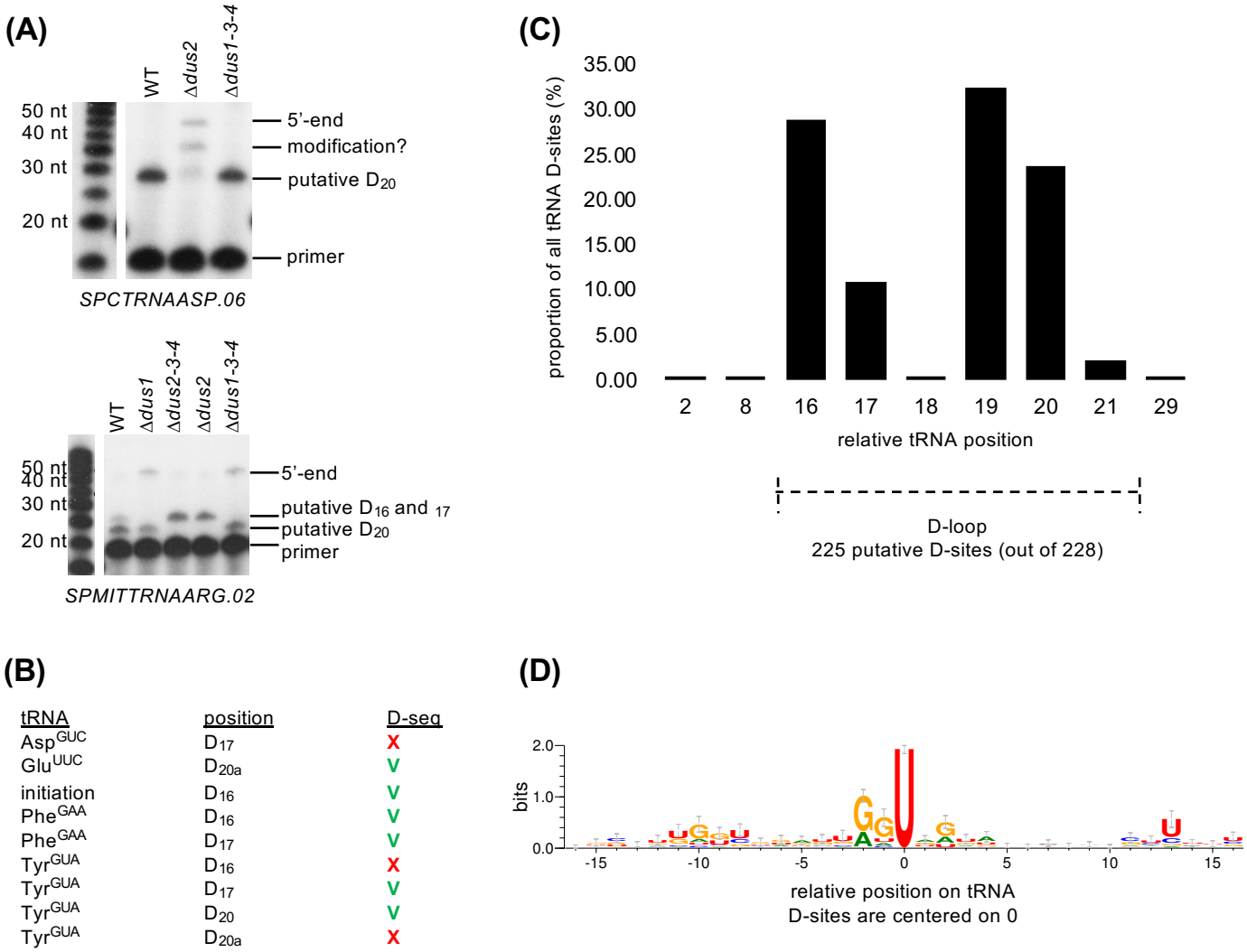


Figure 14 – D-seq detects D on yeast tRNAs. (A) Primer extension assays performed on R+ total RNAs from WT, single *dus* (Δ*dus1*, Δ*dus2*) and triple *dus* (Δ*dus1-3-4*, Δ*dus2-3-4*) mutant strains. D-seq analysis suggested U_{20} modification on *SPCTRNAASP.06* and $U_{16,\ 17}$ and $_{20}$ modifications on *SPMITTRNAARG.02,* that are all confirmed here. **Above**; primer specific for $tRNA^{AspGUC}$, **below**; primer specific for mitochondrial $tRNA^{ArgUCG}$. **(B)** List of known D positions and their cognate tRNAs in *S. pombe*. Last column indicates the detection of the corresponding position (green V) or not (red cross) in the D-seq analysis. **(C)** Distribution of the 228 tRNA D-sites according to their relative position on tRNA. Unlike in Fig. S9, the D-sites were not manually curated to be assigned to a canonical D-site ($D_{16,\ 17,\ 20,\ 20a}$ or $_{47}$). **(D)** Consensus sequence surrounding the 228 tRNA D-sites.

As a proof-of-concept, D-ratios and -fc are shown for $tRNA^{ArgUCU}$ (**Fig. 15**). In WT R+ sample, the D-ratio evolution across the tRNA can be dissected into five regions (**Fig. 15A**); (i) near the 5'-end, a sharp increase of D-ratio was attributed to the natural termination at the end of the transcript, (ii) two close peaks around positions 8-9-10, (iii) a striking peak around nucleotides 16-17, (iv) a more modest increase of the D-ratio at position 21 and (v) a flat region from nucleotide 30 to the 3'-end, corresponding to a zone virtually free of RT termination events. In the control conditions (WT R-; *Δ4* R+ and *Δ4* R-), regions (iii) and (iv) were absent, suggesting that they were *dus*- and labeling-dependent (Fig. 15A). By focusing on region 7-22, it is clear that the positions 8, 9 and 10 all had a D-ratio above 10% across all conditions (**Fig. 15B**). Based on the conservation of the eukaryotic $tRNA^{Arg}$ modifications, the position 9 could be an m^1G (a naturally RT blocking modification) and the position 10 could be an m^2G (a naturally RT pausing modification), explaining the increase of the D-ratio in a *dus*- and labeling-independent manner. Burst of the D-ratio at the position 17, specific to the WT R+ condition, suggested the presence of D_{16}. Moreover, the position 16 itself was characterized by an out-of-background peak both in WT R+ and R-, supporting the idea of a dihydrouridylated residue at this position that naturally paused the RTase. The last peak at position 21 suggested the dihydrouridylation of position 20. Finally, D-fc values above the threshold of 2 were found exclusively on positions 16, 17 and 21 (**Fig. 15C**). Altogether, these profiles implied that $tRNA^{ArgUCU}$ is dihydrouridylated on positions 16 and 20. Implementation of the D-seq pipeline (Fig. 12) gave the same conclusion for the *SPCTRNAARG.08* gene ($tRNA^{ArgUCU}$).

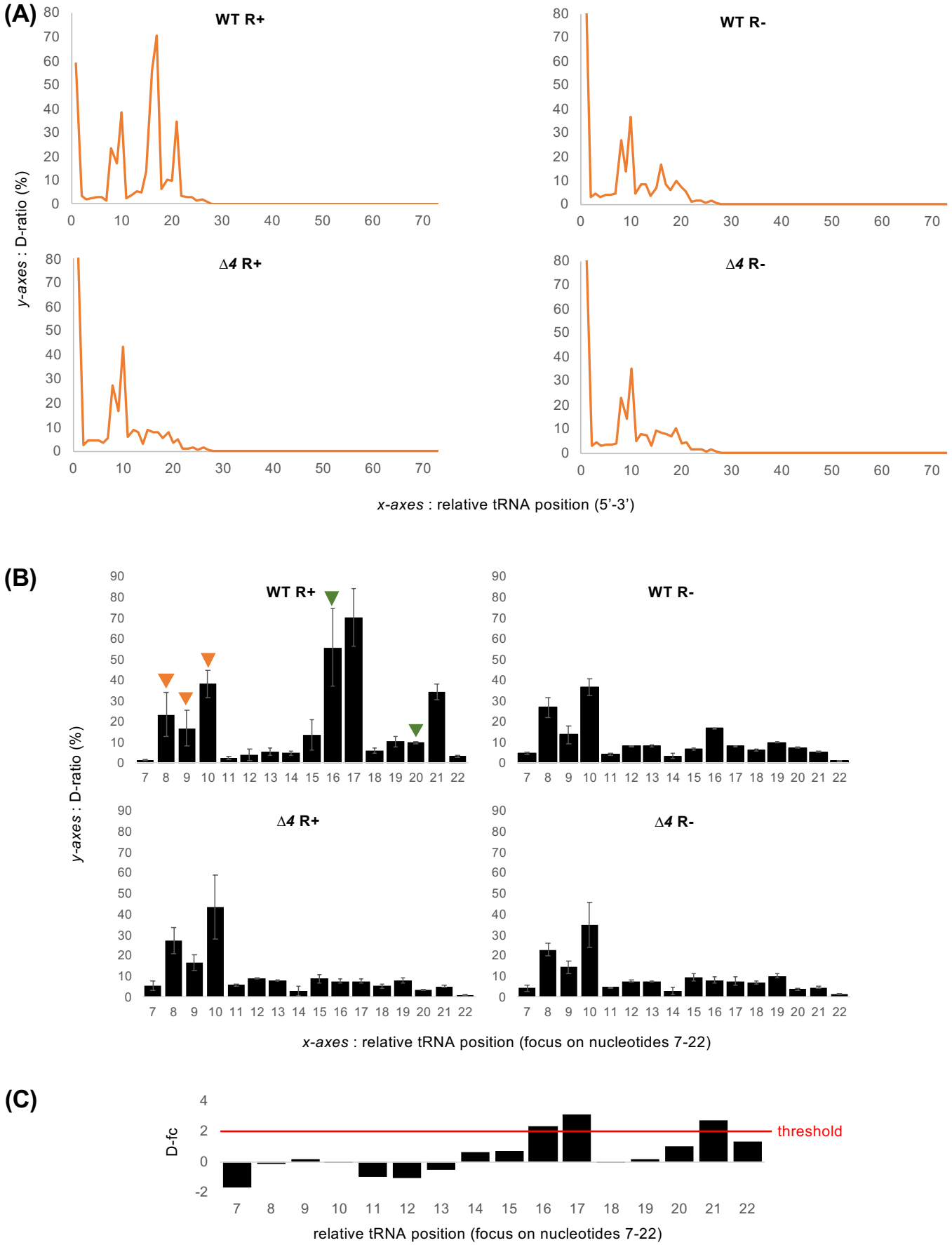


Figure 15 – D-seq is insensitive to RT termination caused by non-D modifications; proof-of-concept with yeast tRNA^{ArgUCU}. (A-B) D-ratio profiles along the entire tRNA (A) or on a zoomed view (B) in test and control conditions. In (B), green triangles represent D-sites (D_{16} and D_{20}) as determined by D-seq, orange triangles represent strong RT termination events that are not taken into account by the analytical pipeline, error bars represent min and max values from two biological replicates. **(C)** D-fold change (D-fc) variation on positions 7 to 22 of *SPCTRNAARG.08*. The red line highlights the D-fc threshold used in the D-seq analysis (Fig. 12).

No D residue was detected on rRNAs, which was in line with the fact that D was never described as a eukaryotic rRNA modification. We used a ribodepletion kit that is not highly efficient on yeast RNA, which permitted to conclude that the detection of D by D-seq was not strictly coverage-dependent (**Fig. S11**). In parallel, we used rRNA sequences to confirm that the RTase was obviously blocked at other transcriptomic positions (**Fig. 16**). The profile of 18S-rRNA in WT R+ exhibited three D-ratio peaks (representing RT terminations); (i) probably a structure-dependent arrest, (ii) a peak corresponding to the naturally blocking $m^1acp^3\psi_{1208}$ RNA modification and (iii) a peak depicting the m^7G_{1616} modification (**Fig. 16A**). Remarkably, the $m^1acp^3\psi_{1208}$ peak was invariable across all conditions (WT R+, WT R-, *Δ4* R+ and *Δ4* R-) in contrast to the situation around m^7G_{1616} (**Fig. 16B**). The example shown here demonstrates that the RT termination also occurred; (i) in a *dus*- and labeling-independent manner (*e.g.* $m^1acp^3\psi_{1208}$) and (ii) in a *dus*-independent manner (*e.g.* m^7G_{1616} that is labeled by Rho). These positions were not taken into account in the D-seq analysis.

In conclusion, the high prevalence of D on expected positions of tRNAs and absence from rRNAs confirm that D-seq is a reliable tool to detect D at a single-nucleotide resolution.

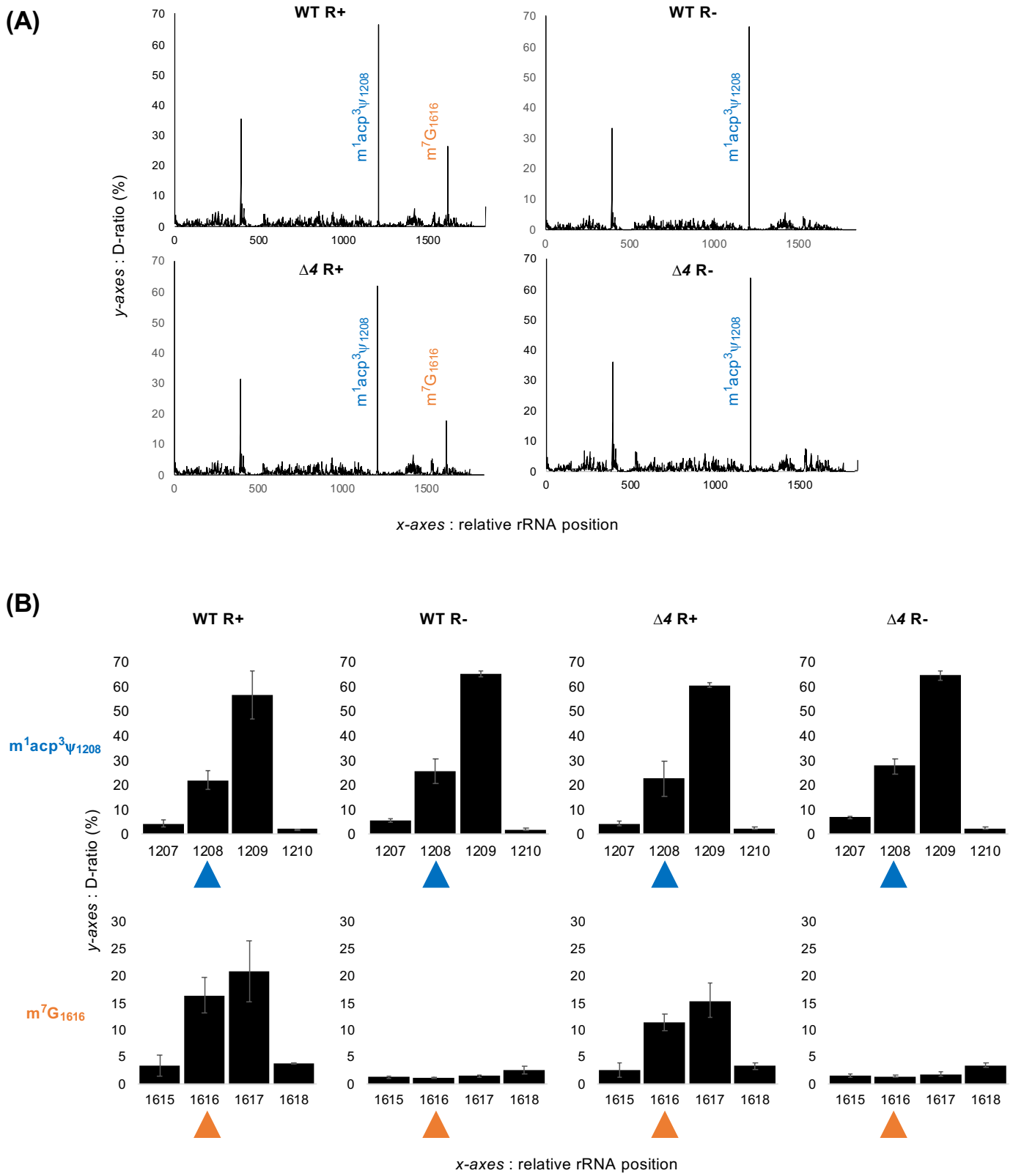


Figure 16 – D-seq is insensitive to RT termination caused by non-D modifications; proof-of-concept with yeast 18S rRNA. (A) D-ratio profiles along the entire 18S rRNA in test and control conditions. The $m^1acp^3\psi$ modification (blue) naturally blocks a reverse transcription reaction. The m^7G modification (orange) is sensitive to Rho labeling and therefore blocked the RTase in R+ conditions. Representative example for *SPRRNA.43*. **(B)** Zoomed view on both 18S-$m^1acp^3\psi_{1208}$ (**above**, the modified site is indicated by a blue triangle) and 18S-m^7G_{1616} (**below**, the modified position is indicated by an orange triangle), error bars represent min and max values from three 18S rRNAs (*SPRRNA.43*, *.44* and *.46*) in two biological replicates.

Yeast coding sequences are dihydrouridylated.

Almost 40% of all 372 detected D-sites were not on tRNAs (**Fig. 17A**). Interestingly, we detected 143 D-sites spread across 125 protein-coding genes (pcg) and 1 site on a lncRNA (**Tables S1** and **S2**). However, this latter overlaps – at the genomic – level with a protein-coding gene that is putatively dihydrouridylated. Most of the pcg (87.2%) had a unique putative dihydrouridine on their sequence and only two genes (*fhn1* and *ala1*) were found to be modified on at least three distinct positions (**Fig. 17B**). D-seq was however not able to discriminate between single mRNAs and it is therefore not possible to know whether *fhn1* and *ala1* carried the three D on the same molecule. Out of the 125 pcg with D, 91 (72.8%) are conserved in vertebrates whereas less than 1% is specific of the *Schizosaccharomyces* genus, suggesting that dihydrouridylation occurs on mRNAs with conserved functions (**Fig. 17C**). Analysis of gene list (AnGeLi, see Experimental procedures) did not reveal any obvious enriched molecular function or transcript feature for the 125 pcg with D (**Fig. S12**). We then implemented a permutation test in which the conditions were randomly mixed before the D-seq analysis (Fig. 12) and compared the number of generated D-sites. As expected, the 16 permutation tests gave on average less than 12 D-sites, which was far less than the 372 detected D-sites when test and control conditions were correctly assigned (**Fig. 17D**).

Another critical parameter is the stoichiometry of dihydrouridylation. The D-seq pipeline does not allow to determine the degree of modification. Notwithstanding this lack of information, we compared D-ratio distributions between D found on tRNAs or mRNAs and observed a decrease for the latter (**Fig. 17E**), suggesting that mRNAs are less heavily dihydrouridylated than their tRNA counterparts.

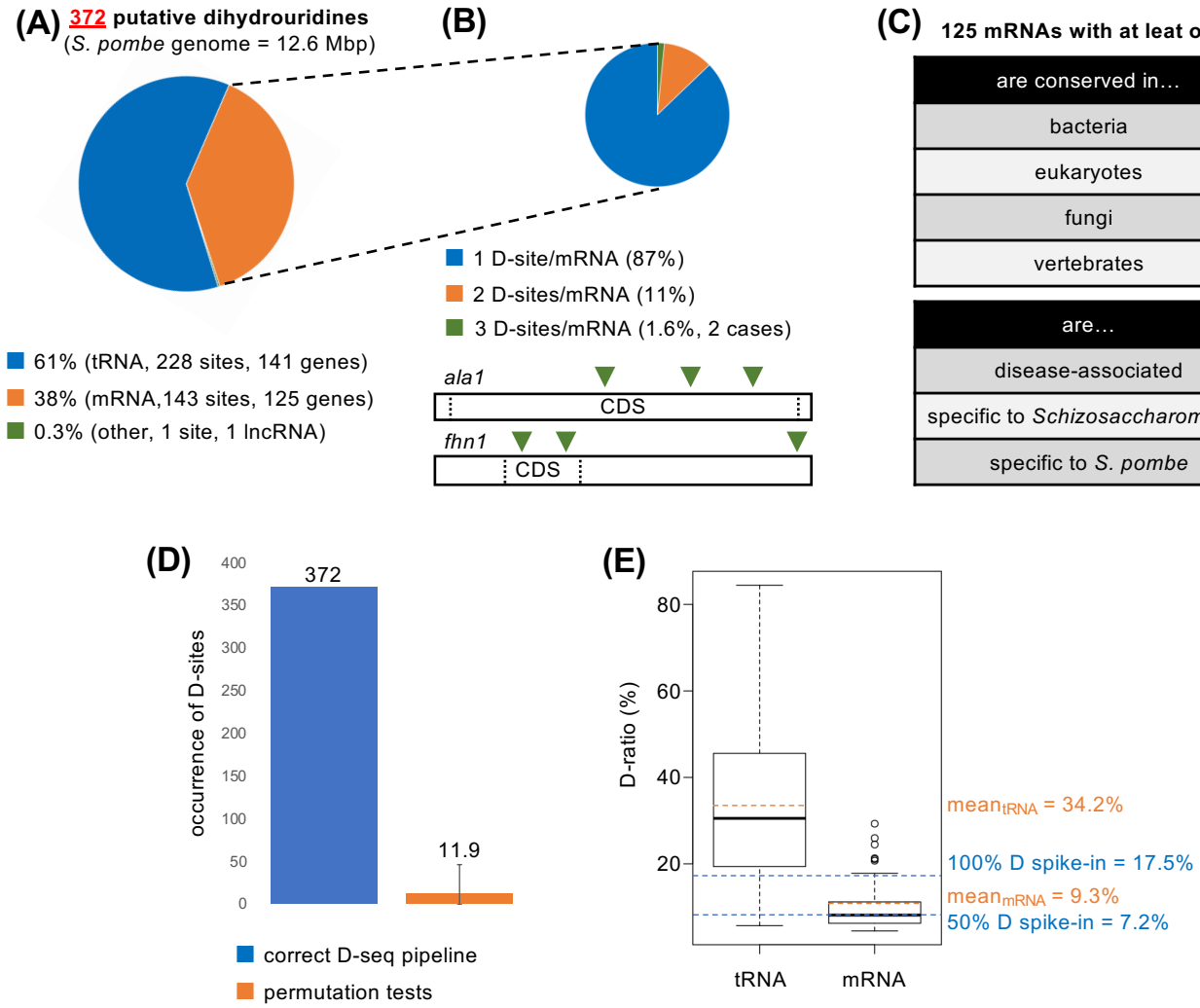


Figure 17 – Yeast protein-coding genes products are dihydrouridylated. (A) Repartition of 372 detected D-sites in fission yeast on three distinct RNA species; transfer, messenger and long noncoding RNAs. **(B)** Putatively dihydrouridylated mRNAs had up to three distinct D-sites on their sequence. *ala1* (coding for an alanine-tRNA ligase) and *fhn1* (coding for an eisosome assembly factor) are the two mRNAs dihydrouridylated at three different positions (green triangles). The CDS (coding DNA sequences) are delimited with black dashed lines. **(C)** Diverse features of the 125 protein coding-genes whose RNA products were modified. **(D)** Correct assignment of test and control conditions was essential in the D-seq analysis pipeline. Sixteen permutation tests were performed in which the properly committed information *1_WT R+; 2_WT R-; 3_Δ4 R+; 4_Δ4 R-* (blue bar, 372 D-sites) was randomly mixed into, for example, *1_WT R-; 2_ Δ4 R-; 3_Δ4 R+; 4_WT R+* (orange bar, average of 11.9 D-sites, error bars represent max and min numbers of detected D-sites). **(E)** Box plot representation of WT R+ D-ratios distribution for tRNAs (228 values) and mRNAs (143 values) and corresponding means highlighted by orange dashed lines (note: this is different from the mean of the D-ratio as presented in Fig. 12, which summarizes the information from test and control conditions). Blue dashed lines indicate the D-ratios of R+ $D_{100\%}$ spike-in (Fig. 13A) and $D_{50\%}$ spike-in (Fig. 13F).

In *S. pombe*, the coding sequence accounts – on average – for 65.0% of the mRNA sequence (including 5'- and 3'-UTRs, CDS and introns). Similarly, 62.4% of all U nucleotides of mRNAs are found within the coding sequence. We observed a sharp enrichment of D on coding sequences with almost 90% of the detected D-sites on CDS (**Fig. 18A**). Consensus analysis of the 124 sequences with a D localized in a coding region revealed a slight propensity for the 5'-DNNNAC-3' sequence (**Fig. 18B**). Interestingly, D seems to be mostly found within the context of a U-tract in the CDS (**Fig. 18C**). In agreement with this observation, the most represented D-containing codon was UUU (**Fig. 18D**). Among the 124 D-sites found on CDS, the most represented codon coded for Leu (**Fig. 18E**). Leucine is coded by six different codons and three of them contain two uridines. For these latter (CUU, UUA and UUG), one could suggest that if dihydrouridylation is random, D would be found in 50:50 ratio on both U. As shown in **Fig. 18F**, distribution of D on Leu codons CUU, UUA and UUG was uneven, suggesting an active mechanism of modification. Moreover, 33% of D-sites (41/124) found on CDS were localized on a codon containing a single U whereas the probability to find D on a single U-containing codon at the transcriptomic level raised to 42% (see Experimental procedures), indicating that the dihydrouridylation preferentially occurs on codons with two or three uridines (**Fig. 18G**). Finally, no D-site was detected on ten different single U-containing codons even though they represent 8% of the coding uridines, and thus account for 8% of the possibilities to get a D-site on the yeast coding transcriptome (**Fig. S13**).

The non-random distribution of dihydrouridine across the transcriptome – and more specifically on products of conserved genes – argues in favor of a biologically relevant dihydrouridylation.

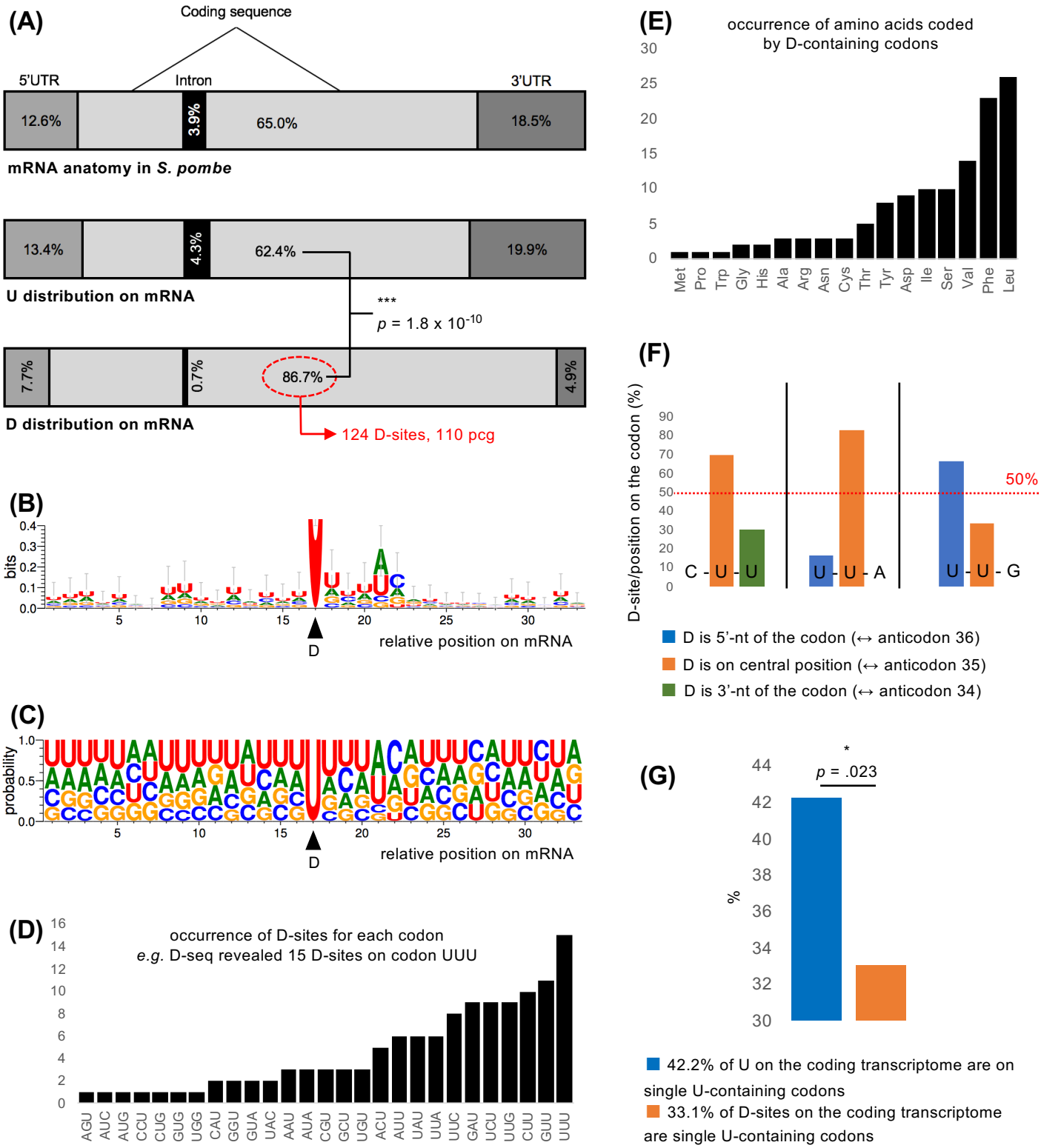


Figure 18 – The yeast coding transcriptome is dihydrouridylated. (A) Enrichment of dihydrouridylation on the coding sequence of mRNAs. **Above**; when a mRNA is 100nt-long, the coding sequence is on average 65nt-long in *S. pombe*, **middle**; when there are 100 U residues on a *S. pombe* mRNA, about 62 of them are on the coding sequence, **below**; 124 (out of 143) detected D-sites were localized on the CDS. pcg (protein-coding gene). One-sided Fisher's Exact test; *** $p < 0.001$. **(B)** Consensus sequence surrounding the 124 coding D-sites. **(C)** Residue probabilities around the 124 coding D-sites. **(D)** The 124 coding D-sites were found on 27 different codons ranging from one occurrence (for AGD, ADC, ADG, CCD, CDG and DGG) to fifteen (for DUU, UDU and UUD). **(E)** The 27 represented codons of Fig. 18D code for 17 different amino acids. **(F)** Non-random distribution of D-sites on Leu codons containing two U residues. **(G)** Under-representation of D-sites on codons containing a single U. One-sided Fisher's Exact test; * $p < 0.05$.

Absence of dihydrouridine affects tRNA but not mRNA or protein abundances.

We compared the transcriptomic profiles of WT and Δ*4* strains by RNA-seq to assess the role of D on RNA stability. Only a limited set of genes were differentially expressed (|fc| > 1.5, FDR < 0.1) (**Fig 19A**) and no overlap between these genes and the putatively dihydrouridylated protein-coding genes was observed. On the contrary, hypodihydrouridylated tRNAs were affected by the loss of D and were generally more abundant, although not always significantly (**Fig. 19B**). We confirmed this pattern for the D_{16}/D_{17}-containing $tRNA^{PheGAA}$ by Northern blotting, but not for $tRNA^{CysGCA}$ for which no D-site was isolated by D-seq (**Fig. S14**). The transcriptomic analysis of a dihydrouridine-free strain thus provided experimental evidence to show that D is critical to control global tRNA abundance.

Among the 125 dihydrouridylated mRNAs, three out of eight mRNAs encoding subunits of the CCT chaperonin complex carried potential dihydrouridines (*cct1*, *cct6* and *cct7*). The CCT complex is required for the co-translational folding of a set of proteins, including tubulins α (*nda2*) and β (*nda3*), whose mRNAs were also putatively dihydrouridylated. qRT-PCR analyses suggested a slight decrease – albeit not significant (see legend of Fig. 19) – of *nda2* and *cct1* mRNAs relative abundances in a strain lacking D. Other mRNAs that were not D-seq candidates displayed the same profile; the control RBP *nab3* as well as the CCT targets *slp1* (non-cytoskeletal) and *act1* (cytoskeletal). The 25S rRNA control was unaffected by the loss of D (**Fig. 19C**). Surprisingly, Nda2 and Cct4 proteins were somehow more abundant in the Δ*4* strain (**Fig. 19D**).

The transcriptomic analysis and the assessment of protein production in the quadruple *dus* mutant revealed that D does not strikingly impact the putatively dihydrouridylated mRNAs and therefore support the idea of a subtle role played by the modification.

(A)

logFC	gene name	associated function
-3.86	*dus1*	tRNA dihydrouridine synthase Dus1
-2.95	*dus2*	tRNA dihydrouridine synthase Dus2
-1.73	*prp4*	serine/threonine protein kinase Prp4
-1.56	*dus3*	tRNA dihydrouridine synthase Dus3
-1.50	*mfm2*	M-factor precursor Mfm2
-0.99	*SPAC27D7.09c*	But2 family protein
-0.79	*SPBC1348.02*	S. pombe specific 5Tm protein family
-0.79	*SPBPB2B2.19c*	S. pombe specific 5Tm protein family
-0.77	*SPAC977.01*	S. pombe specific 5Tm protein family
-0.76	*SPAC750.05c*	S. pombe specific 5Tm protein family
-0.76	*SPNCRNA.1071*	intergenic RNA
-0.66	*hsp9*	heat shock protein Hsp9

logFC	gene name	associated function
0.65	*SPMITTRNAGLU.01*	tRNA Glutamic acid, mitochondrial
0.69	*ura2*	dihydroorotase Ura2
0.78	*lcp1*	cyclin L family cyclin
0.85	*SPATRNAGLU.01*	tRNA Glutamic acid
0.85	*SPBTRNAGLU.06*	tRNA Glutamic acid
0.85	*SPBTRNAGLU.07*	tRNA Glutamic acid
0.85	*SPATRNAGLU.02*	tRNA Glutamic acid
1.30	*SPBTRNAILE.07*	tRNA Isoleucine
1.33	*SPCTRNAVAL.09*	tRNA Valine
1.40	*SPATRNACYS.01*	tRNA Cysteine
1.52	*SPATRNAILE.04*	tRNA Isoleucine
1.53	*SPATRNACYS.03*	tRNA Cysteine
2.06	*SPATRNACYS.02*	tRNA Cysteine
4.97	*SPNCRNA.1212*	antisense RNA (predicted)

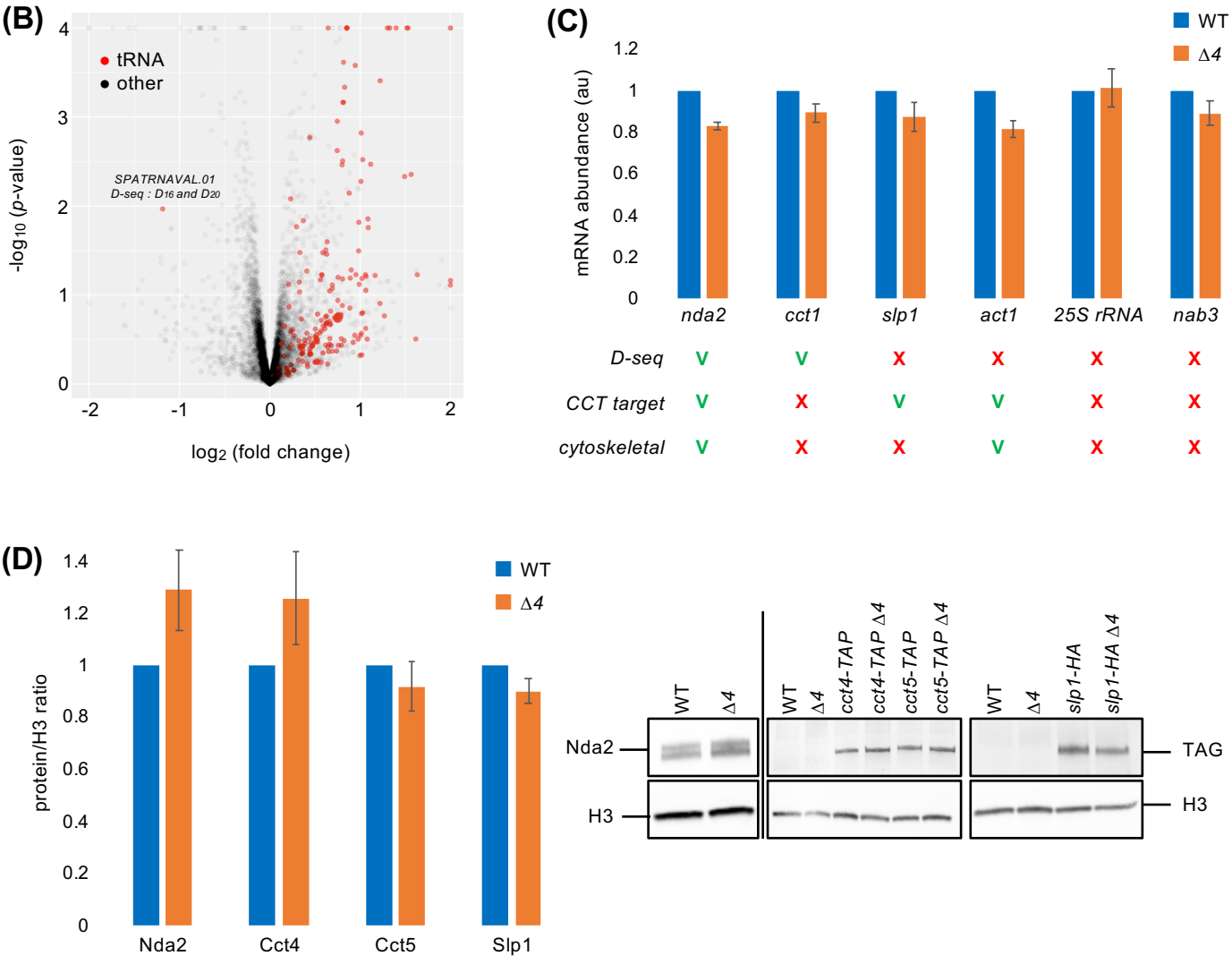


Figure 19 – Transcriptomic and protein analyses of the Δ*4* mutant. (A) Significantly and differentially expressed genes revealed by RNA-seq on WT and Δ*4* total RNA samples. **On the left,** list of downregulated genes in the quadruple *dus* mutant. **On the right,** list of upregulated genes in the quadruple *dus* mutant. logFC stands for $\log_2$ (fold change). **(B)** Volcano plot highlighting all tRNA genes (red dots) and all other fission yeast transcripts (black dots) mapped in the RNA-seq experiment. **(C)** qRT-PCR analyses ($\Delta\Delta C_T$ method) on D-seq candidates whose protein products are targeted by the CCT folding complex (green V) or not (red cross), are cytoskeletal proteins (green V) or not (red cross). *25S rRNA* is a noncoding transcript control, *nab3* is a coding transcript control, WT expression values (blue) are scaled to 1 (au, arbitrary units), Δ*4* expression mean values are in orange, *n* = 3, error bars represent SEM, normalization is done on *adh1* expression. **(D)** Relative protein expression of one D-seq candidate (Nda2), two tagged CCT subunits (Cct4 and 5) and one target of the CCT complex (Slp1). Histone H3 is used as a loading control. **On the left,** ratio (protein of interest/H3) is scaled at 1 for WT (blue) and variations are shown for the Δ*4* mutant (orange), *n* = 3, error bars represent SEM. **On the right,** representative result of a Western blotting triplicate against Nda2 (upper left, anti-Nda2 Ab – antibody), Cct4/5-TAP (upper middle, anti-TAP Ab), Slp1-HA (upper right, anti-HA Ab) and H3 (lower, anti-H3 Ab) in different genetic backgrounds. For C and D, Student's *t*-tests were performed (on the Cq values or on the ratio (protein of interest/H3), respectively) and differences between WT and mutant were all non-significant ($p > 0.05$).

Dihydrouridylation is limited to tRNAs in fast-growing *E. coli* cells.

Because 30% of the fission yeast dihydrouridylated mRNAs were conserved in bacteria (Fig. 17C), we performed D-seq on fast-growing *E. coli* cells. We worked with the previously generated triple *dus* mutant (Δ*dusA-B-C* or Δ*3*) (Bishop et al., 2002). Dot blot and Northern blotting analyses of R+ and R- treated RNAs led to the same conclusion drawn for yeast; overall decrease of Rho signal in the Δ*3* genetic background (**Fig. 20A** and **B**).

Details of the *E. coli* tRNA modifications are available in databases – unlike fission yeast tRNAs – providing a thorough set of information to compare D-seq hits and previously known D positions. We detected 106 D-sites on tRNA positions 16 to 23 of 62 different genes. Manual curation (using tRNAmodviz and GtRNAdb) of each predicted D-site revealed 63 distinct modified nucleotides on 35 different tRNA species. Strikingly, 53 were expected D residues, representing 72% of all known dihydrouridylated tRNA positions in *E. coli*. Furthermore, we propose ten new dihydrouridylation events on six tRNA species (**Fig. 20C**). In total, the 63 D-tRNAs are distributed along the D-loop by following the exact same pattern of the 74 known D (**Fig. 20D**). Altogether, these data confirm the reliability of D-seq to detect D at the single-nucleotide level.

Not a single D-site was detected on mRNA or rRNA. The 23S-U_{2449} position is nevertheless known to be dihydrouridylated in *E. coli*. Interestingly, RT termination signals were observed around this position in both WT R+ and Δ*3* R+, but not in mock-treated samples (**Fig. 20E**), as confirmed with a primer extension assay (**Fig. 20F**). In other words, RTase was stopped one nucleotide downstream of D_{2449} after Rho labeling but in a *dusA-B-C*-independent manner, strongly suggesting that ribosomal dihydrouridine in installed by a *fourth* dihydrouridine synthase. Another possibility is that this position was incorrectly considered as a D, but a recent report confirmed the position 2449 as dihydrouridylated (Popova and Williamson, 2014). It is reasonable to ask whether this putative supplementary Dus enzyme could act on mRNAs. To test this hypothesis, we modified the D-seq statistical analysis (Fig. 12) to get rid of the *dusA-B-C* dependency for the determination of RT termination events; if there are D residues that are not installed by DusA, DusB or DusC, these sites are inevitably *dusA*, *dusB* and *dusC*-independent. To do so, we calculated the *p*-values on the *labeling* factor (R+/-) alone, therefore ignoring the *strain* factor (WT/*dusA-B-C*). In other words, the test conditions encompassed both WT R+ and Δ*3* R+ whereas the control conditions were limited to WT R- and Δ*3* R- conditions. The modified pipeline led to the detection of RT termination events at 40 additional sites including; (i) five positions on different mRNAs (*aldA*, *pheS*, *putA*, *yicJ* and *ynhF*), (ii) 23S-U_{2068} that is upstream of the known 23S-m^7G_{2069}, (iii) several tRNA positions with notably 13 s^4U_8 positions and (iv) two wobble positions ($tRNA^{GlnUUG}$-$cmnm^5s^2U_{34}$ and $tRNA^{LysUUU}$-$mnm^5s^2U_{34}$).

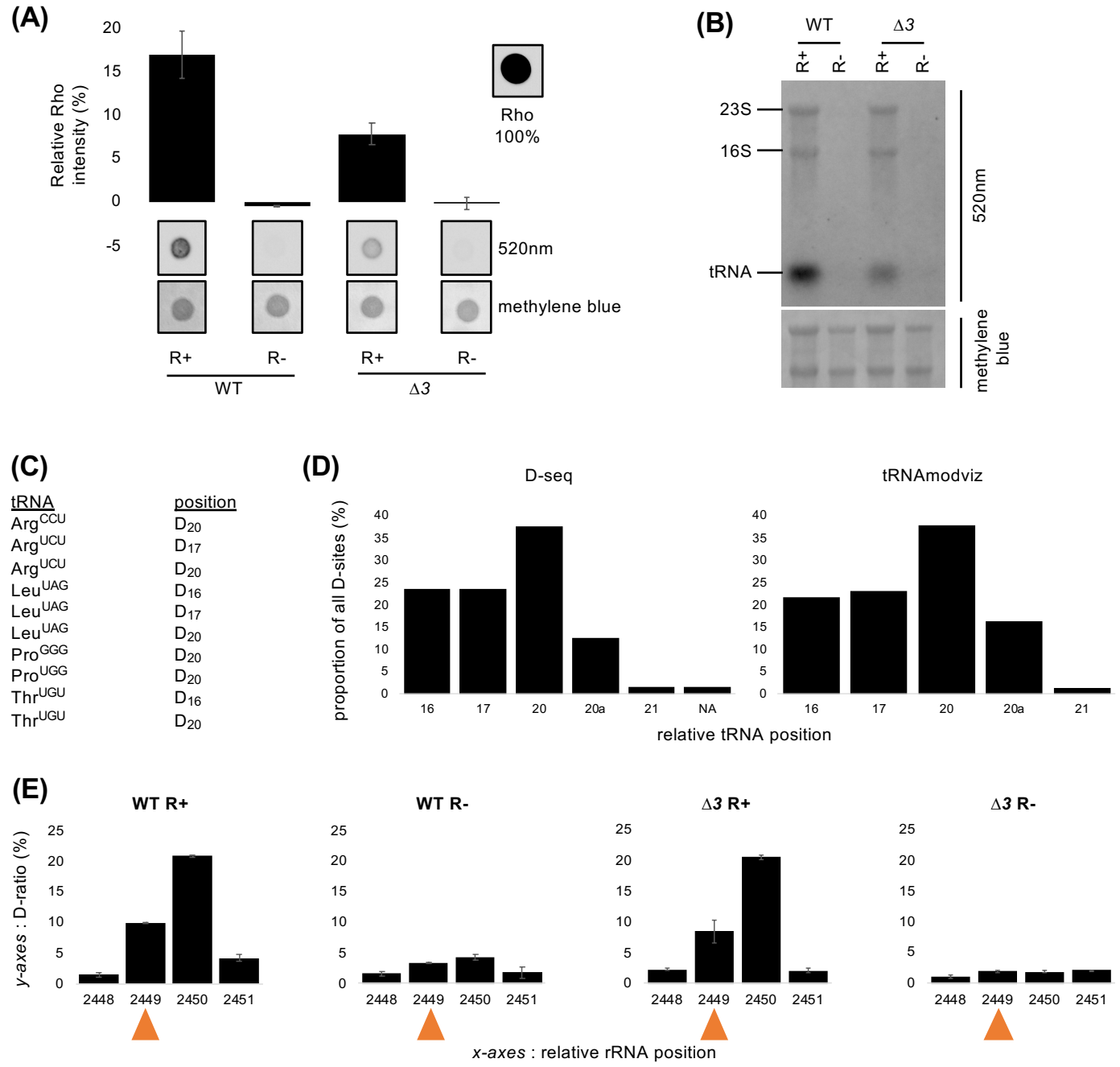


Figure 20 – D-seq implementation on *E. coli* transcriptome. (A) Comparative Rho detection at 520nm on WT (wild type) and Δ*3* (Δ*dusA-B-C*) total RNAs in R+ and mock-treated (R-) samples. The intensity of a Rho drop is scaled at 100%, a drop of water is considered as the background, methylene blue staining serves as loading control, *n* = 3, error bars represent SEM. **(B)** Fluorescent Northern blotting of WT and Δ*3* total RNAs in R+ and mock-treated samples. Representative result from a biological triplicate, methylene blue staining serves as loading control. **(C)** Ten previously unconsidered positions (tRNAmodviz) on six *E. coli* tRNA species are proposed to be dihydrouridylated. **(D) On the left,** distribution of the 63 unique tRNA D-sites according to their relative position on tRNA. D_{20} is the most represented dihydrouridylated position with 24/64 counts (37.5%). NA stands for *not assigned* due to sequence incompatibility (D_{22} on Ser^{GGA}). **On the right,** distribution of the 74 unique tRNA D positions as referred by the tRNAmodviz database. D_{20} is the most represented dihydrouridylated position with 28/74 counts (38%). **(E)** D-ratio (%) profiles around the 23S rRNA-U/D_{2449} position (orange triangles) in test and control conditions. Similarities between WT and Δ*3* R+ profiles suggest a DusA-B-C-independent synthesis of D_{2449}. D-ratio values were obtained by the compilation of the seven *E. coli* 23S rRNA genes (*rrlA*, *B*, *C*, *D*, *E*, *G* and *H*) from two biological replicates, error bars represent min and max values of the D-ratio means from both replicates. **(F)** Primer extension assay performed on total RNAs from WT and Δ*3* strains, R+ or R- treated. The radiolabeled primer targeted the 23S rRNA. Representative result of a biological triplicate.

D is undetectable in strains lacking all predicted *dus* genes.

As introduced hereabove (see Introduction III.C.i.b. and III.C.ii.b.), screens for bacterial and eukaryotic Dus enzymes did not rely on methods that could undeniably recover all genes/proteins coding for/carrying a dihydrouridine synthase activity. HPLC analysis was performed on total RNA extracts from WT and Δ*4* yeast strains (**Fig. 21A**) and from WT and Δ*3* bacterial strains (**Fig. 21B**). In both cases, no residual D signal was observed and proper detection of D was controlled by calibration with a commercial D (**Fig. S15**).

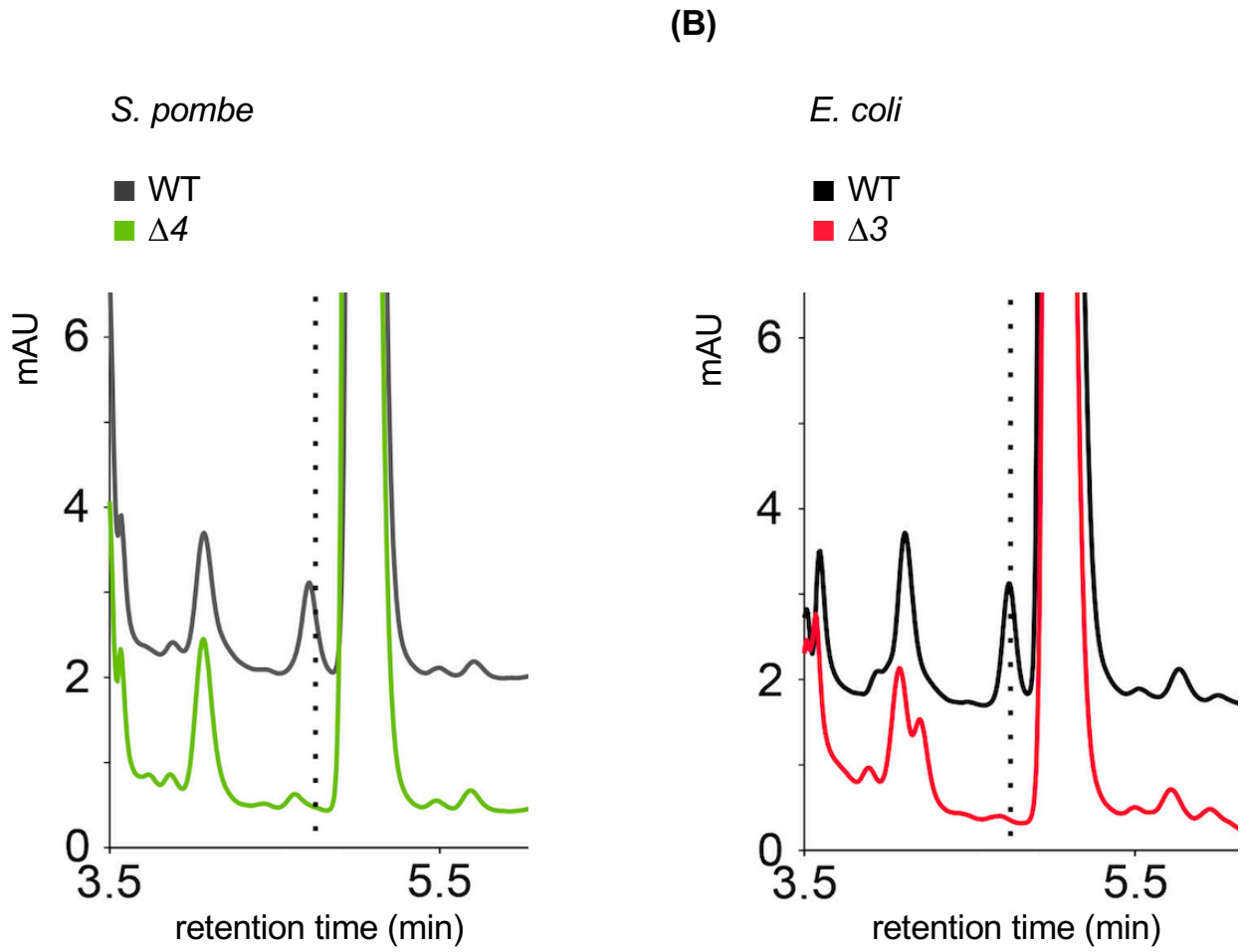


Figure 21 – D is virtually undetectable in yeast and bacterial *dus* mutants. Analysis of total RNA from *S. pombe* **(A)** and *E. coli* **(B)** WT and *dus* mutant strains by HPLC for three biological samples. Retention time for D was determined with commercial dihydrouridine (Fig. S15) and is indicated by a black dotted line. Detection at 247nm. mAU (milli-absorbance units).

Growth of mutants lacking D is affected by the temperature

In the light of the correlation between tRNA dihydrouridylation and growth temperature (see Introduction III.C.ii.c.), we tested the growth of the yeast Δ*4* mutant at different temperatures and observed that the viability was impaired at high but not low temperatures (**Fig. 22A**). Importantly, we showed that the *ts* phenotype was not caused by the underexpression of the gene upstream of *dus4* in the Δ*4* genetic background (**Fig. S16**), which was shown to be less abundant in the quadruple mutant (Fig. 19A). Remarkably, yeasts grown at 20°C had significantly more dihydrouridine than the ones grown at the optimal temperature (32°C) (**Fig. 22B**).

The growth phenotype of the Δ*3* bacterial mutant was also affected by the environmental temperature (**Fig. 22C**). At 37°C, the mutant displayed a slight growth delay that was hardly aggravated at 42°C. At low temperatures, the *dus* mutant had an obvious growth defect compared to the WT strain.

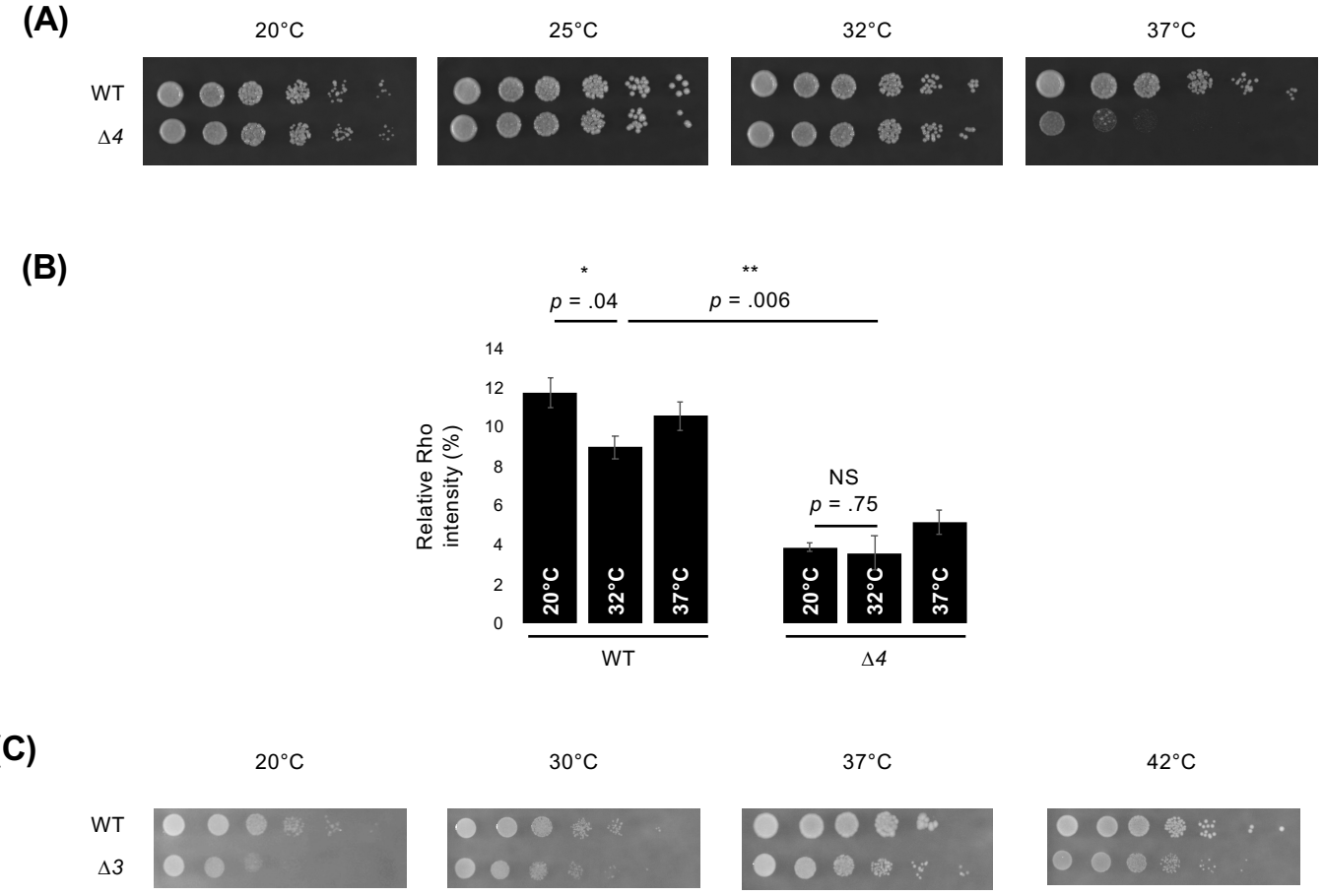


Figure 22 – Correlation between D and temperature. (A) Growth phenotype of the quadruple Δ*4* yeast mutant compared to the WT strain. Exponential growing cells were spotted on rich medium and incubated at different temperatures for 48 (32 and 37°C), 90 (25°C) or 115 hours (20°C). Representative growth from three biological replicates. **(B)** Comparative Rho detection at 520nm on WT and Δ*4* yeast total RNAs in R+ samples. RNA was harvested from yeasts grown to mid-exponential phase at 20, 32 or 37°C. The intensity of a Rho drop is scaled at 100%, a drop of water is considered as the background, *n* = 3, error bars represent SEM, Student's *t*-test; * $p < 0.05$, ** $p < 0.01$, NS $p > 0.05$. **(C)** Growth phenotype of the triple Δ*3 E. coli* mutant compared to the WT strain. Exponential growing cells were spotted on rich medium and incubated overnight at different temperatures (30, 37 and 42°C) or for 48 hours (20°C). Representative growth from three biological replicates.

Dihydrouridine is not essential in human cells.

Within the framework of this project, *dus* mutants were also generated in a human cell line (HCT 116 cells producing functional p53) by using the CRISPR-Cas9 technology (**Fig. S17**). The strategy relied on the electroporation of two ribonucleoprotein complexes (RNPs) per targeted gene. Each RNP was constituted by the Cas9 endonuclease and two noncoding RNAs (the sequence-specific crRNA and its transactivating partner). The first coding exon of the gene of interest (*DUS1*, *DUS2*, *DUS3* or *DUS4*) was specifically targeted at two distinct positions – separated by a few hundred nucleotides –, therefore necessitating the electroporation of two RNPs, each of them inducing a DNA double-strand break. Single homozygous mutants for the four *DUS* genes were obtained and confirmed by sequencing (**Table 4**, see Experimental procedures). Strikingly, the specificities of human DUS enzymes for tRNA substrates seemed to be conserved (Fig. 7A), as suggested by primer extension assays on different tRNA targets (**Fig. 23A**). In parallel, HPLC profiles suggested that all four enzymes did not contribute equally to the total pool of dihydrouridines with each DUS protein supporting ~ 9 to 25% of total D levels (**Fig. 23B**).

In an attempt to get a quadruple *DUS* mutant (*Hs*Δ*4*), eight RNPs were electroporated in the same pool of cells (2 RNPs x 4 *DUS* genes). Because HCT 116 cells are diploid, sixteen events of double-strand breakdowns were expected to occur to get a homozygous quadruple mutant. The procedure was efficient and resulted in the generation of a viable *Hs*Δ*4* mutant, as confirmed by sequencing. Importantly, the *Hs*Δ*4* mutant had virtually no D (**Fig. 23C**), and therefore demonstrated that loss of D is not lethal, even in higher eukaryotes.

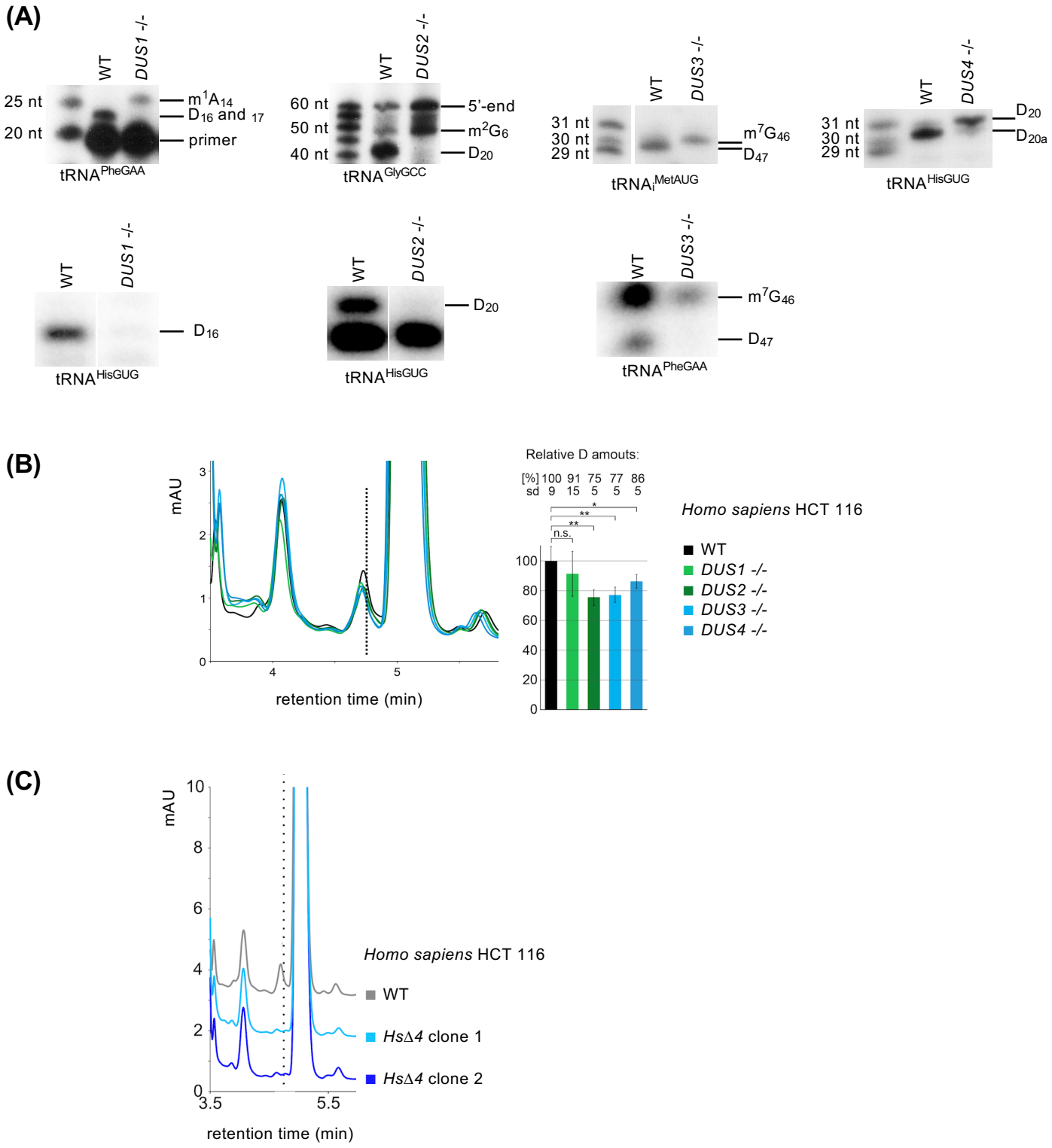


Figure 23 – Characterization of human *dus* mutants. (A) Primer extension assays performed on total RNAs from HCT 116 isogenic cell line and homozygous single *DUS* mutants (-/-, an exon is partially or entirely deleted, see Fig. S17). The targeted tRNA is indicated below each panel. **Above**; R+ treated RNAs, **below**; aniline-based protocol to induce RT termination. **(B-C)** Analysis of total RNA from HCT 116 WT and *DUS* mutant cell lines by HPLC. Retention time for D was determined with a commercial dihydrouridine (Fig. S15) and is indicated by a black dotted line. (B) Respective contribution of the four DUS enzymes to D synthesis in total RNA. *n* = 3, Student's *t*-test; * $p < 0.05$, ** $p < 0.01$, n.s. $p > 0.05$. sd (standard deviation). (C) HPLC analysis of D content in two quadruple *DUS* mutants (see Table 4). For B and C, detection at 247nm, mAU (milli-absorbance units).

Absence of phenotype in the Δ*4* mutant facing different stresses.

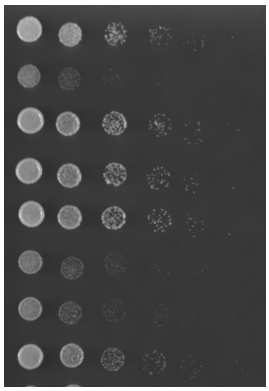
In an attempt to explore the biological relevance of mRNA dihydrouridylation, we tested the growth of the Δ*4* mutant in the presence of numerous chemical compounds that have various targets (**Fig. 24A**). Strikingly, the *dus* mutant was much more sensitive than the WT strain to the tubulin destabilizing compound TBZ (**Fig. 24B**). However, the phenotype was the consequence of the *dus4* upstream gene *prp4* differential expression in a Δ*4* background (Fig. 19A and Fig. S16). In parallel, we also investigated the spindle formation in a *dus* mutant at 37°C (**Fig. 25A**). The dynamics was not significantly different between WT and mutant strains, but a D-deprived strain turned out to be smaller at 37°C (**Fig. 25B**). Importantly, we worked with a Δ*dus1* Δ*dus2* Δ*dus3* $dus4_{C108A\text{-}K149A}$ mutant (Cys_{108} and Lys_{149} are important for enzymatic activity) to avoid pleiotropic effects of the *dus4* deletion.

(A)

conditions	***comments***
0.1% glucose + 2% galactose (+ adenine)	respiration medium (~~fermentation~~)
0.1% glucose + 3% glycerol	respiration medium (~~fermentation~~)
defined medium (EMM)	free of adenine, uracil, leucine, histidine, *etc.*
cold shock 10, 30 or 60min at 4°C	/
heat shock 10, 30 or 60min at 42°C	/
heat shock 30min at 40°C	/

chemical compounds	***comments***
antimycin A	bocks e^- transfer from cytochrome b to cytochrome c_1
myxothiazol	prevents e^- transfer from Fe-S center to ubiquinone
rotenone	prevents e^- transfer from Fe-S center to ubiquinone
oligomycin	inhibits F_0 ATP synthase
methothrexate	inhibitor of folic acid and purine metabolisms
thiabendazole*	tubulin-destabilizing compound
cycloheximide	inhibits translation elongation, disrupts cytoskeleton
sodium chloride*	infers actin defects and osmolarity remediability
latrunculin A	prevents actin polymerization
benomyl*	binds and inhibits MTs
nocodazole	insensitivity of WT yeasts, inhibits MT polymerization
hydrogen peroxide	oxidative stress, known to increase aggregates in yeast
2-azetidine carboxylic acid	prevents folding of newly synthesized proteins
trivalent arsenic*	prevents tubulin polymerization, inhibits CCT complex

(B) + TBZ 12.5µg/mL



WT
Δ*4*
WT + empty vector
WT + *dus4*
WT + *prp4*
Δ*4* + empty vector
Δ*4* + *dus4*
Δ*4* + *prp4*

Figure 24 – Phenotyping the yeast Δ*4* mutant. (A) Sensitivity of the quadruple *S. pombe* Δ*4* mutant was assessed by spot assay and compared to the WT growth in all listed conditions. Asteriks indicate that the mutant was more sensitive than the WT strain but most likely in *dus*-independent manner, as exemplified in (B). MT (microtubule). **(B)** Growth phenotype of the Δ*4* mutant compared to the WT strain. Exponential growing cells were spotted on rich medium supplemented with 12.5µg/mL of thiabendazole (TBZ) and incubated at 32°C (5-fold dilutions). All strains carried an auxotrophic marker for the positive selection of the vector during (pre-)culturing growth in a defined medium. The vector used in this study resulted in the overexpression of the cloned gene (*dus4* or *prp4*).

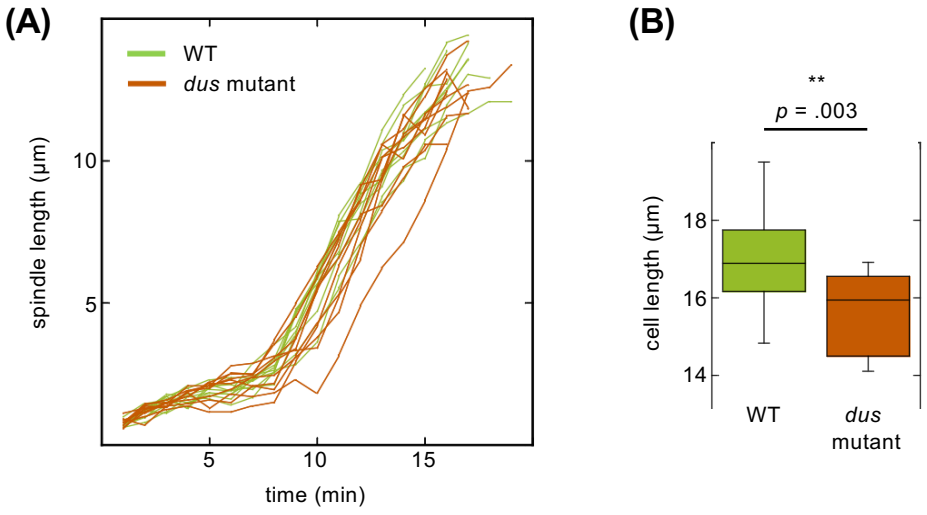


Figure 25 – *S. pombe* Δ*dus1* Δ*dus2* Δ*dus3* *dus4*$_{C108A-K149A}$ mutant spindle dynamics and cell length. (A) Spindle dynamics is assessed by measuring the length of the spindle throughout the cell cycle in WT (green) and *dus* mutant (orange) strains at 37°C. Microtubules are fluorescently tagged (*mCherry-tubulin α2*). **(B)** Δ*dus1* Δ*dus2* Δ*dus3* *dus4*$_{C108A-K149A}$ cells (genotypically different of a Δ*4* mutant) are smaller than WT cells at 37°C. *n* = 20, ** p < 0.01.

ADDENDUM

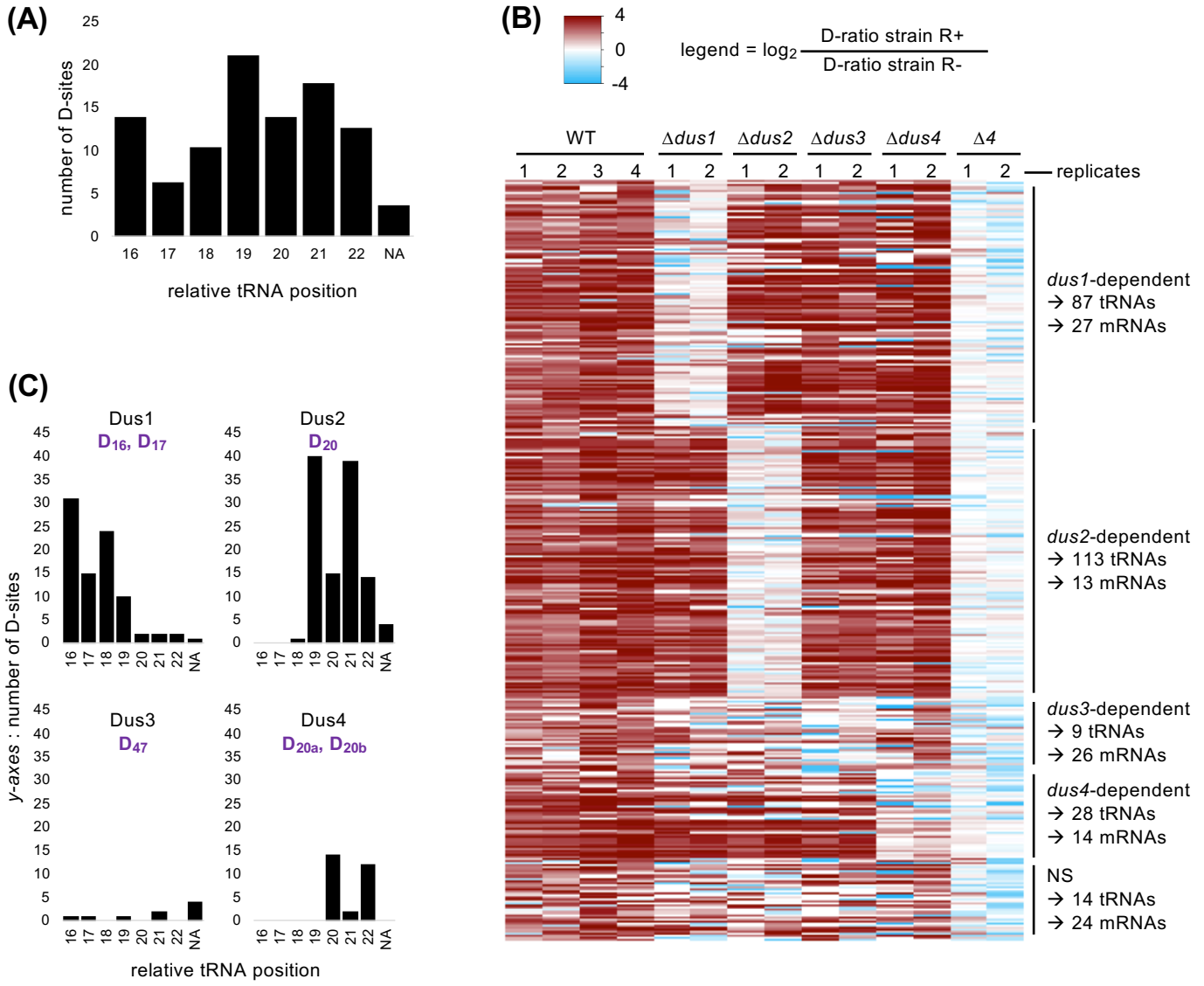
Within the course of the preparation of this dissertation, new results were obtained about the *dus* dependency of the detected D-sites in *S. pombe*. These new sets of data have not been fully analyzed and interpreted yet, explaining why they are found in this special section.

At the experimental level, total RNAs were extracted from five different strains (WT and the four single *dus* mutant; Δ*dus1*, Δ*dus2*, Δ*dus3* and Δ*dus4*), subsequently labeled (R+ and R-) and prepared for sequencing (Fig. 11B). The analytical pipeline was similar to the one shown in Fig. 12, expect that it relied on the comparison of four biological replicates for the WT conditions and two for the Δ*4* conditions. For each detected D-site, the *dus* dependency was then determined according to the FDR of the statistical test (< 10%, Fig. 12). For example, the D-site *X* was considered as *dus1*-dependent if the statistical test on the Δ*dus1* strain generated an FDR < 0.1 (and FDRs > 0.1 for the tests based on the Δ*dus2*, Δ*dus3* and Δ*dus4* strains).

At the tRNA level, 251 D-sites were detected and 96.4% of them were located between positions 16 and 22 (**Fig. 26A**), as it could be expected (Fig. 7A, Fig. S9). Interestingly, 87, 113, 9 and 28 D-sites were considered as *dus1*-, *dus2*-, *dus3*- and *dus4*-dependent, respectively (**Fig. 26B**). For the remaining sites (14, representing 5.6% of the 251 detected D-sites on tRNAs), the *dus* dependency could not statistically be determined. The single-nucleotide distribution of the D-sites did not perfectly match with the substrate specificities of the Dus enzymes (Fig. 7A) (**Fig. 26C**). This can be explained; (i) by the necessity to implement a manual curation of all sites (a predicted D_{18} could actually correspond to D_{17} [Dus1] or D_{19} [Dus2]) and (ii) by working on the definition of *dus*-dependency. Indeed, there are many cases where the FDR is below the threshold (0.1) not only for one single mutant, but for several or even all of them (examples are given in **Fig. 26D**). In these cases, the *dus*-dependency was determined according to the smallest FDR when several FDRs were below the threshold of 10%. This pipeline is not ideal but is convenient for a preliminary analysis and will be thoroughly improved for further investigations.

In parallel, D-sites were also detected on protein-coding genes; 104 modifications on 92 mRNAs. When compared to the Fig. 17, 20 D-sites were found in both analyses (**Table S3**). Importantly, 90% of the D-sites (94/104) were found on a coding region (**Fig. 26E**) and U-rich codons were the most represented (**Fig. 26F**), thus confirming our conclusions about the yeast coding transcriptome dihydrouridylation. Fig. 26B depicts the *dus*-dependency of the D-sites found on mRNAs, although the analytical pipeline and thresholds have still to be refined.

These data – even though they are the result of an exploratory and still ongoing analysis – highlighted a good correlation between the lack of a Dus enzyme and the detection of dihydrouridines on tRNAs. This observation seemed to be more easily monitored for *dus1*- and *dus2*-dependent modifications (Fig. 26B). However, it appeared possible that the deposition of D at a specific tRNA position could not be strictly dependent of one Dus enzyme (Fig. 26D). The example of *SPBTRNAASN.02* suggested that the D_{20} deposition might not be exclusively *dus2*- (and so Dus2-)



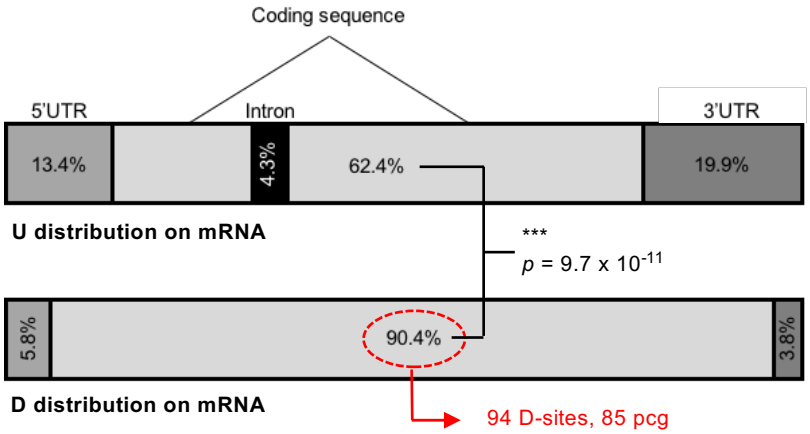
(D)

tRNA gene	relative position	manually curated position	*dus1*	*dus2*	*dus3*	*dus4*	comment
SPATRNAGLU.04	16	16	8.9×10^{-236}	1	1	1	Dus1-D
SPATRNAGLU.04	20	20a	1	1	1	0	Dus4-D
SPBTRNAGLN.04	19	20	1.7×10^{-152}	0	9.1×10^{-13}	5.6×10^{-31}	probably Dus2-D but not clear
SPBTRNAGLN.04	20	20a	0	5.3×10^{-112}	0.5	0	effect of D_{20} ?
SPBTRNAASN.02	20	20	3.9×10^{-4}	9.6×10^{-18}	1	1	probably Dus2-D but not clear
SPBTRNAASN.02	21	20a	0.2	2.0×10^{-11}	0.4	1.53×10^{-18}	effect of D_{20} ?

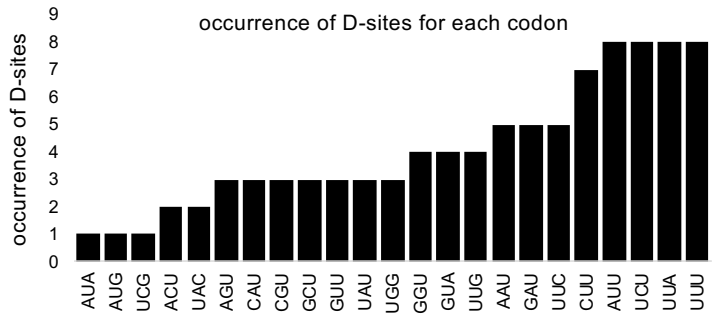
corrected *p*-value (FDR)

Figure 26 →

(E)



(F)



→ **Figure 26 – Preliminary results from the investigation on *dus*-dependency. (A)** Distribution of the 251 tRNA D-sites according to their relative position on tRNA. NA stands for *not assigned,* for positions that were out of the D-loop. **(B)** Heatmap showing all detected D-sites that were sorted by comparison of their respective D-ratios in test (R+) and control (R-) conditions. The dark red hue highlights more RT terminations in the test condition (D-ratio R+ > D-ratio R-). The columns depict all replicates from six different strains (WT, four single *dus* mutant and the quadruple *dus* mutant). All sites were clustered according to their *dus1*-, *dus2*-, *dus3*- or *dus4*-dependency, as shown to the right of the map. If no dependency was revealed by the statistical analysis, the sites were clustered in the NS (*non statistical*) category. **(C)** Distribution of the 87 (Dus1-dependent), 113 (Dus2-dependent), 9 (Dus3-dependent) and 28 (Dus4-dependent) tRNA D-sites according to their relative position on tRNA. NA stands for *not assigned,* for positions that were out of the D-loop. The expected modified sites (Fig. 7A) are shown in purple. **(D)** Examples of *dus*- (and so Dus-) dependency determinations for three tRNA genes (column 1). Columns 2 and 3 indicate the relative or manually assigned tRNA position that is considered as a D-site. FDR values resulting from the statistical analyses on the *dus*-dependency are displayed in columns 4 to 7 (FDR < 0.1 as a threshold). *SPATRNAGLU.04* is unambiguously decorated by a Dus1-dependent D_{16} and a Dus4-dependent D_{20a}. **(E)** Enrichment of dihydrouridylation on the coding sequence of mRNAs. **Above**; when there are 100 U residues on a *S. pombe* mRNA, about 62 of them are on the coding sequence, **below**; 94 (out of 104) detected D-sites are localized on the CDS. pcg (protein-coding gene). One-sided Fisher's Exact test; *** $p < 0.001$. **(F)** The 94 coding D-sites are found on 23 different codons.

dependent and that the D_{20a} installation might require Dus2 or a Dus2-dependent dihydrouridylation, explaining why the deletion of *dus2* resulted in the loss of D_{20a} both in *dus2* and *dus4* mutants. Deepening the analysis will certainly allow to consider the potential enzymatic plurality as a biological truth or as a technical/*in silico* artifact. Furthermore, the patterns observed for mRNAs depicted; (i) that multiplying biological replicates of the test condition (WT R+) did invariably lead to the detection of D on coding regions of the yeast transcriptome and (ii) that there is apparently not a single Dus enzyme that is responsible for the modification of mRNAs in yeast.

DISCUSSION AND PERSPECTIVES

Working on dihydrouridine; a wise choice?

Since the description of dihydrouridine as a naturally occurring RNA modification (Holley et al., 1965; Madison and Holley, 1965), molecular biologists have tried to unravel the functionality of this unique modification. It is unique because it is the sole known nonaromatic modification, feature that promotes noncanonical ribose conformation and hinders proper base stacking. Unlike the well-studied 2'-O-methylation and pseudouridine RNA modifications that are described as enhancers of regional stability in ribonucleic acids, D installation is considered as a blueprint of conformational flexibility (Dalluge et al., 1996b). In parallel, *in silico* and structural analyses shed on light on the phylogenetic conservation of Dus enzymes and on the molecular tricks they acquired for substrate specificities (Bou-Nader et al., 2018; Byrne et al., 2015; Kasprzak et al., 2012; Xing et al., 2004).

The singularity of D modification makes it an interesting example to unravel how the structural nature of ribonucleic acids is modulated using a simple, yet unique modification. All these reasons explain why investigations focusing on dihydrouridine are of high significance.

Rhodamine labeling; a reliable tool?

In this study, the dihydrouridine mapping at the single-nucleotide resolution is achieved through the binding of a stable fluorophore at the modified residue and subsequent RT termination one nucleotide downstream. Combination of the fluorescent labeling of D with primer extension assay had never been tested before but turned out to be a very successful technique.

The first questionable aspect is the reduction of dihydrouridine by sodium borohydride (Fig. 5 and 6); is it selective and is the unmodified uridine resistant to this chemical? The uracil nucleobase is characterized by a reactive C6 involved in the C5=C6 double bond that is part of the electrophilic Michael acceptor. In dihydrouridine, C5=C6 is lost and the resulting nucleobase stability is affected in reducing conditions (i.e. $NaBH_4$). Sodium borohydride is undoubtedly selective for dihydrouridine as confirmed and published for 50 years (Betteridge et al., 2007; Cerutti et al., 1968a; Cerutti et al., 1968b; Cerutti and Miller, 1967; Igo-Kemenes and Zachau, 1969; Kaur et al., 2011; Molinaro et al., 1968). A uridine can actually be reduced by $NaBH_4$ into dihydrouridine but it requires very specific conditions (light-dependent reaction at 50°C with excess of $NaBH_4$ in a basic solution) (Cerutti et al., 1968b). Our data are in agreement with the selective reduction of dihydrouridine in the tested conditions with subsequent covalent binding of rhodamine. Indeed, a reaction of reverse transcription was never blocked at an unmodified U position – as verified with synthetic and $\Delta 4$ RNAs (Fig. 10C-F, 13C-D, S6 and 14A).

The second legitimate question is about the selectivity of the chemical reaction for dihydrouridine compared to other RNA modifications. The answer is clear; D is not the only posttranscriptional modification to be sensitive to sodium borohydride and/or to be labeled with rhodamine; many reports indeed established the sensitivity of s^4U, ac^4C, m^7G and yW (Beltchev and Grunberg-Manago, 1970; Cerutti et al., 1968a; Igo-Kemenes and Zachau, 1969; Molinaro et al., 1968; Schleich et al., 1978; Wintermeyer et al., 1979; Wintermeyer and Zachau, 1971, 1974, 1979). Dot blot and fluorescent Northern blotting were in line with these observations since the loss of (virtually) all D residues in *Δ4* and *Δ3* mutants did not result in a complete drop of rhodamine detection (Fig. 10A-B and 20A-B). Moreover, the sequencing data after R+ treatment clearly indicate the sensitivities of m^7G (Fig. 16B) or s^4U (*E. coli* D-seq) for rhodamine labeling. In other words, the rhodamine labeling is not specific for dihydrouridine and it is thus reasonable to conclude that other labeled RNA modifications induce RT arrests. That is the reason why the *dus* mutants are not only powerful controls, but essential conditions. In all our experiments – including D-seq – the *dus* mutants served as a background control to ensure that the WT R+ signal was indeed *dus*-dependent. It also means that our transcriptome-wide pipeline could easily be transferred to other Rho labeling-sensitive modifications; *thiI* is a nonessential gene in *E. coli* and is responsible for s^4U tRNA thiolation (Mueller et al., 2001), Trm8 is a tRNA-methyltransferase whose deletion results in loss of m^7G in budding yeast tRNAs (Alexandrov et al., 2002), *etc.* (**Table 3**).

Finally, another issue is to know what precise molecular mechanism leads to the rhodamine incorporation. As already mentioned previously (see Introduction III.B.iii. and Fig. 6), there are two potential processes following the sodium borohydride reduction; the formation of a THU intermediate or the opening of the D-ring. Without experimental evidence, it is currently impossible to assume one model more than another but it seems that technical details such as $NaBH_4$:uridine ratio, pH and temperature are critical for reactivity (Betteridge et al., 2007; Kaur et al., 2011; Pan et al., 2009). Moreover, it is important to highlight that we applied the Rho labeling on total RNA samples whereas the aforementioned studies worked solely with purchased or *in vitro* prepared tRNAs, which makes the biochemical environment completely different.

In conclusion, despite the lack of evidence about the precise mechanism behind the Rho labeling of D residues, the technique is undeniably reliable to detect D residues at a single-nucleotide resolution. *dus* mutants are essential controls to discriminate between D and other RNA modifications. Biased (primer extension) and unbiased (D-seq) approaches both confirm selectivity of the Rho labeling for dihydrouridines and not uridines.

D-seq; trustworthy modification calling?

The pipeline used to assign a nucleotide as a dihydrouridine is an improved version of the two-metric system imagined by Regev and colleagues for pseudouridine mapping in ψ-seq (Schwartz et al., 2014). We changed both the technical and analytical aspects of the pipeline used to detect putative ψ-sites, in order to make D-seq more stringent. First, the calculation of the mean D-ratio constituted a powerful pre-filtering step for three main reasons (Fig. 12); (i) it was independent of the statistical test (Bourgon et al., 2010), (ii) it was unbiased *sensu stricto* because it did not compare

RNA modification	chemical reactivity	*E.coli*	*S. cerevisiae*	*S. pombe*	*H. sapiens*	references
ac^4C	$NaBH_4$ not in the tRNA structure	TmcA	Kre33	Nat10	NAT10	(Cerutti et al., 1968a; Cerutti and Miller, 1967; Igo-Kemenes and Zachau, 1969, Molinaro et al., 1968)
f^6A	$NaBH_4$ not clear				FTO	(Cerutti and Miller, 1967)
m^1A	reversible reactivity with $NaBH_4$ not in the tRNA structure		Gcd10 Gcd14 Bmt2 Rrp8	Gcd10 Gcd14 Bmt2 Rrp8	TRMT6 TRMT61A TRMT61B TRMT10C (mt) BMT2 RRP8	(Cerutti et al., 1968a; Igo-Kemenes and Zachau, 1969; Macon and Wolfenden, 1968)
m^3C	$NaBH_4$		Abp140	Trm140 Trm141	METTL2 METTL6 METTL8	(Cerutti et al., 1968a; Macon and Wolfenden, 1968)
m^7G	$NaBH_4$ and dye labeling	RlmL RsmG TrmN	Abd1 Bud23 Trm8 (Trm82)	Pcm1 Bud23 Trm8 (Trm82)	RNMT BUD23 METTL1 WDR4	(Igo-Kemenes and Zachau, 1969; Wintermeyer and Zachau, 1970, 1971, 1974) suggested by D-seq
s^2U	R+ dependent RT arrest of s^2U derivatives (*e.g.* $mnm^5s^2U_{34}$)	MnmA	Ncs6 Ncs2 Slm3 (mt)	Ctu1 Ctu2 Slm3 (mt)	CTU1 CTU2 TRMU (mt)	suggested by D-seq
s^4U	$NaBH_4$	ThiI				(Cerutti et al., 1968a; Cerutti and Miller, 1967; Igo-Kemenes and Zachau, 1969) suggested by D-seq
yW	$NaBH_4$ and dye labeling		Tyw1 Trm12 Tyw3 Ppm2	Tyw1 Trm12 Tyw3 Ppm2	TYW1 TRMT12 TYW3 LCMT2	(Beltchev and Grunberg-Manago, 1970; Igo-Kemenes and Zachau, 1969; Schleich et al., 1978; Wintermeyer and Zachau, 1970, 1971, 1974, 1979) suggested by D-seq

Table 3 – Chemical reactivity of different RNA modifications for sodium borohydride ($NaBH_4$) and dye labeling. The first column (RNA modification) encompasses all the RNA modifications that are susceptible to be used in a D-seq-like pipeline, although the chemical reactivity shown in the second column (chemical reactivitiy) does not necessarily generate an RT termination. The four next columns give the associated known or predicted RNA modifying enzymes in four species (*Escherichia coli*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and *Homo sapiens*). Trm82 are non-catalytic proteins. (mt) means that the enzyme has mitochondrial substrates. In yeasts, *abd1*, *pcm1*, *kre33*, *S. cerevisiae gcd10* and *gcd14*, *S. pombe slm3* are essential genes. yW formation also implies the Trm5 methyltransferase that is not indicated here. The last column (references) depicts all studies that showed the chemical reactivity/ies of the corresponding RNA modifications.

conditions to one another, and (iii) it did not imply the usage of an arbitrary threshold because the mean D-ratio threshold was based on the stoichiometry of the spike-in RNA (Fig. 13F). Moreover, D-seq was performed with two biological replicates, which permitted the implementation of a statistical test (Fig. 12). It means that all detected D-sites were relevant because the downstream nucleotide – or RT-stop site – was statistically more represented as being the last retro-transcribed residue in the test condition. Finally, the control condition was not limited to a mock-treated sample (R-) but extended to the use of a *dus* mutant that unarguably allows to distinguish the Dus-dependent RT terminations. The filtering steps established before and after the statistical analysis ensured a proper modification calling.

It is also important to highlight that the possible loss of a transcript in the *dus* mutant does not lead to a false positive site calling; if a transcript *T* is decorated with a modification *M*, if *M* is labeled by the rhodamine and thus generates an RT termination and if *T* is absent in the mutant – for example because the deletion of the *dus* genes impacts its expression and/or abundance, the D-ratio for *M* will be high in WT R+ and low in all other conditions. However, the low D-ratio in the *dus* mutant would be an artifact simply because the transcript *T* is absent. The statistical analysis will nevertheless not call the *M* position because the RT termination event occurring around *M* will not be *statistically* relevant. This example demonstrates how important is the implementation of a statistical test.

There are only two scenarios for which a site would be wrongly called (false positive); (i) the RT termination randomly occurs more in WT R+ in both biological replicates but not in other conditions (which is intrinsically not impossible but the probability is low) and (ii) the different thresholds used in the analytical pipeline are too permissive. This latter is unlikely because we determined the threshold values to be as stringent as possible. Our purpose was to minimize the false positive rate, even though it increased the miss rate (or false negative rate). As a consequence, D-seq is not exhaustive at all but instead relies on a high confidence modification calling. As a proof-of-concept, 98.5% of all detected D-sites (225/228 in yeast and 104/106 in *E. coli*) were found on canonical D positions. The detection of remaining sites is discussed in **Fig. 27**. The modification calling of the only D of the synthetic RNA was also in line with proper determination of D-sites (Fig. 13A). Furthermore, not a single D-site was detected on yeast r-, sn- and sno-RNAs, as expected (Fig. S11) (Cantara et al., 2011; Gu et al., 1996). However, it is important to keep in mind that positive and negative controls might have two drawbacks when intended to be compared to the modification calling on mRNAs; (i) t- and r-RNAs are highly expressed, decreasing the signal:noise ratio compared to the coding transcripts and (ii) the synthetic RNA is not dihydrouridylated *in vivo* and added to total RNA before Rho labeling, potentially impacting the biochemical identity of this exogeneous dihydrouridine. These minor cons are, though, not sufficient to challenge the trustworthy modification calling applied in D-seq.

Another questionable aspect of the method relies on the detection of neighboring D. In a sequence 5'-D_1D_2N-3', the RT will stop at the N position in the R+ condition, allowing the detection of D_2. But does the D-seq permit the detection of D_1? There are several examples where two neighboring D were detected (*e.g. SPATRNAPHE.01* with D_{16} and D_{17}, *SPATRNAVAL.02* with D_{20} and D_{20a} or the coding *SPAC694.02* gene with a predicted UDD codon). Although it is not possible to known

S. pombe

tRNA	D-site	comment
Arg^{CCU}	D_2	possible 5'-end transcript artifact
Met^{CAU}	D_8	possible 5'-end transcript artifact
Pro^{CGG}	D_{29}	position close to the D-stem, could be a structural-based and D-dependent arrest because a D_{20} modification is detected on the same tRNA

E. coli

tRNA	D-site	comment
Ser^{GGA}, *serX* gene	D_{22}	sequence incompatibility between tRNAmodviz and BioCyc → D_{22} is likely D_{20a}
Ser^{GGA}, *serW* gene	D_{22}	sequence incompatibility between tRNAmodviz and BioCyc → D_{22} is likely D_{20a}

Figure 27 – **Rare unexpected D calling sites on tRNAs.** D-seq reveals less than 2% of unexpected tRNA D-sites that could be explained by simple technical artifacts.

whether the two modified residues were on the same molecule, this observation suggests that the D-seq does indeed detect two following D. In parallel, a primer extension performed on an *in vitro* transcribed RNA with three subsequent D indicates that the RT arrest is more prominent downstream of the third D (with the RNA point-of-view, Fig. S6). It is therefore reasonable to speculate that D-seq should highlight neighboring dihydrouridylated residues but to a lesser extent for the 5'-dihydrouridines (still with the RNA point-of-view).

To sum up, D-seq combines both wet and dry manipulations that turned out to be accurately controlled to draw meaningful conclusions. It is worth reiterating that our conclusions are *not*; (i) D-seq provides a full picture of dihydrouridylated mRNAs in yeast and (ii) there is no D on bacterial mRNAs. Instead, we assume that; (i) D-seq provides a non-exhaustive list of dihydrouridylated mRNA candidates that are dependent on Dus1, 2, 3 and/or 4, and (ii) Dus A, B or C-dependent D were not detected in our conditions in bacterial mRNAs.

Yeast D-seq mRNA candidates; what to think?

About 100 pseudouridines were found onto yeast mRNAs (Carlile et al., 2014; Lovejoy et al., 2014; Schwartz et al., 2014), providing an example of a low abundant internal mRNA modification in a *simple* organism grown on *normal* conditions. ψ profiling was then discussed by Jaffrey: "A central question is whether pseudouridylation in mRNA is biologically meaningful. The apparent lack of a dedicated mRNA pseudouridylase raises the possibility that these pseudouridines reflect nonspecific pseudouridylation" ((Jaffrey, 2014), p. 946). The same concern could (and probably will) raise about dihydrouridylation of yeast mRNAs.

The dihydrouridines spread beyond the tRNA pool could be a random background that is tolerated by the cell because it is not harmful. Despite the lack of experimental evidence, several concepts are important to mention; (i) dihydrouridine is such a peculiar RNA modification (does not stack, promotes RNA flexibility, spans the RNA chain, *etc.*) that it is hard to conceive that D deposition does not influence structural – so functional – identities of RNA molecules. It is established that D is a nonessential modification for *normal* growth but nonessential does not mean non-functional. A salient feature of RNA modification enzymes is that they are not necessarily required for growth (Phizicky and Hopper, 2010; Sharma and Lafontaine, 2015). At the structural level, unmodified and native tRNAs adopt the same overall shape (reviewed in (Phizicky et al., 2009)), relegating tRNA modifications to a second role. In budding yeast, a set of double mutants coding for modifying enzymes are viable at 30 but not 37°C (Alexandrov et al., 2006). For example, the double Δ*trm8* Δ*dus3* mutant (lacking the tRNA-m^7G_{46} and D_{47}) is not viable at 37°C, phenotype that is associated to RTD (see Introduction I.B.iii.a.) (Chernyakov et al., 2008). Even though a unique modification does not often seem to be functionally important, it is critical to consider a set of intertwining parameters that act together to confer a biological role. The same scenario is possible for the dihydrouridylation of eukaryotic mRNAs; dihydrouridine plays a subtle role along with other *cues* (modifications, expression level, environment, *etc.*) that needed to be studied together to unravel the biological role of D; (ii) in the same context, the detected D-sites could be a baseline set of modified residues whose relevance is prominent in a specific physiological context. Two interesting cases are illustrated in the literature; induced pseudouridylation upon

heat shock and widespread methylation during the meiotic program (Schwartz et al., 2013; Schwartz et al., 2014). Due to the prevailing statement that ψ and D have opposite structural functions in RNA (stability *vs* flexibility), it is remarkable that pseudouridylation is induced upon heat shock whereas we observed more D upon cold growth (Fig. 22B). Moreover, because D is suggested to be an adaptive response to low temperature for psychrophilic bacteria (see Introduction III.C.ii.c.), it could be worth performing D-seq on yeast grown at 20°C. In parallel, dot blot assay also revealed a higher signal of Rho at 37°C, but both in the WT and Δ*4* backgrounds (Fig. 22B), suggesting a D-independent increase of the fluorescence. However, because the Δ*4* mutant is thermosensitive (Fig 22A), performing D-seq on yeast grown at 37°C could also bring a valuable set of information; (iii) importantly, conservation of the modification pattern would be a clue that mRNA-D have a biological significance. That is the reason why D-seq is being performed with the mutants generated in a human cell line and data will be soon available. However, the absence of detected D-sites on bacterial coding transcriptome is a first relevant information because it suggests that mRNA dihydrouridylation could be a eukaryotic-specific mechanism. Moreover, it means that Dus enzymes do not seem prone to randomly dihydrouridylate nonspecific substrates, hence a conceivable biological relevance for eukaryotic mRNA D candidates; (iv) finally, the nonrandom distribution of D on mRNAs strongly argues for a targeted dihydrouridylation on coding sequences that underlies an active and regulated mechanism of modification (Fig. 18), especially on products of conserved genes (Fig. 17C).

dus mutants; what phenotype?

The function of dihydrouridine synthases is uncovered – i.e. they catalyze the formation of D – but their role(s) as biological regulators in a physiological context is still unknown. To unravel this issue, we tested the sensitivity of the Δ*4* mutant in various conditions and facing different chemical compounds (Fig. 24). Moreover, a recent study reported *dus2* and *dus3* mutants as phenotypically associated with promiscuous conjugation, incomplete cell disassembly at cell fusion site and decreased mating efficiency in fission yeast (Dudin et al., 2017). All these characteristics are correlated to aspects of developmental biology involving the cytoskeleton, thus leading to the analysis of the spindle formation in a *dus* mutant, which did not display apparent defects in this specific context (Fig. 25). From these observations, we can conclude that the thermosensitivity of the yeast Δ*4* mutant (Fig. 22A) is probably not explained by deficiencies in the biogenesis of microtubules and might be essentially explained by the disruption of hypomodified tRNAs at an elevated temperature, which is line with the RTD phenotypes described above (see Introduction I.B.iii.a.). It is therefore appropriate to point out that the description of a phenotype that would be correlated to the dihydrouridylation of mRNAs is/will be a painful/impossible task because of the tRNAs are always affected in *dus* mutants.

The *E. coli* Δ*3* mutant phenotyping is not a simple matter either because the Δ*3* mutant (Bishop et al., 2002) is actually a quadruple mutant consisting of a Δ*dusA* Δ*dusB* Δ*fis* Δ*dusC* genetic background (**Fig. 28**). *dus* genes were only partially deleted in the Δ*3* mutant but the regions coding for residues involved in catalysis, cofactor binding or tRNA binding were targeted, strongly suggesting that they were catalytically inactive (Bou-Nader et al., 2018; Byrne et al., 2015; Savage et al., 2006; Yu et al., 2011). *dusB* and *fis* are on the same operon. Fis is a transcriptional factor that

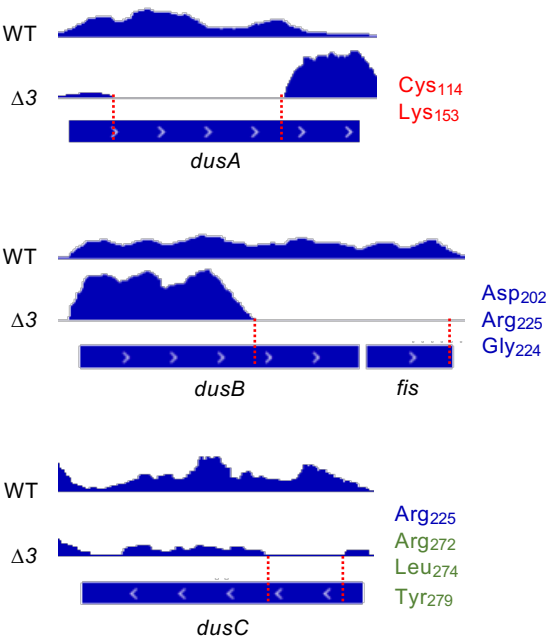


Figure 28 – *dusA, B* and *C* genomic loci in WT and *Δ3* strains. Blue *waves* represent the D-seq unnormalized coverage profiles in WT and *Δ3* R+ conditions, replicate 1. Blue recantgles represent protein-coding genes. **Above:** *dusA* deletion includes residues experimentally shown to be important for catalytic activity, including the highly conserved cysteine (red). **Middle:** *dusB* 3'-end deletion includes residues involved in polar contacts with the $FMNH^-$ proton donor (blue). *fis* is completely deleted. **Below:** *dusC* deletion includes one residue involved in polar contacts with the $FMNH^-$ proton donor (blue) and three residues interacting with the tRNA (green; Arg_{272} makes a salt bridge with the tRNA-G_{19}-U_{20} phosphodiester bond, Leu_{274} interacts with the tRNA-C_{56}, Tyr_{279} participates in recognition of the tRNA-U_{16}-C_{17}) (Bou-Nader et al., 2018; Byrne et al., 2015; Savage et al., 2006; Yu et al., 2011).

(in)directly regulates one fifth of the transcriptome (Cho et al., 2008). Importantly, WT and Δ*3* expression profiles are not compared to each other in the D-seq experiment. It means that the *fis* deletion should not impact the D-site calling, except if the loss of Fis-dependent transcriptional activation results in the loss of a set of transcripts. However, the absence of detectable D-sites on the bacterial mRNAs might be caused by the large-scale transcriptional disruption provoked by the absence of Fis. It is thus not possible to undeniably conclude about the non-dihydrouridylation of *E. coli* mRNAs. At the regulatory level, *dusB-fis* is a tightly regulated operon with different DNA binding sites for Fis itself, the cAMP receptor protein Crp and the DNA supercoiling protein IHF (Caspi et al., 2016). In bacteria, operons are often organized in clusters of functionally-related genes (Ralston, 2008). The fact that *dusB* and *fis* are under the control of the same genomic elements could give insights about the regulation of *dusB* and its potential physiological role. Interestingly, *dusA* could also be differentially expressed, more specifically upon heat shock because it has a temperature-dependent σ^{32} consensus binding site (Zhao et al., 2005). Furthermore, the sensitivity of the Δ*3* mutant at 20°C (Fig. 22C) is somehow interesting for further investigations but needs to be confirmed as *fis*-independent.

Bridging the epitranscriptome to the phenotype is a challenging step. Moreover, it happens that the full deletion of an RNA modification enzyme gene results in a phenotype that is less severe in the corresponding catalytic mutant (Agarwala et al., 2012). It suggests the importance of the protein itself – not only the catalytic activity, therefore complicating the dissection between modification-dependent and modification enzyme-dependent phenotypes. The molecular determinants for Dus activity are known and conserved (Kasprzak et al., 2012; Savage et al., 2006), enabling the creation of Dus catalytic mutants for proper phenotyping in Bacteria and Eukarya[4]. A last aspect to keep in mind is that the nonessentiality of D modification could imply a compensation mechanism in which the cell deals with the loss of D by reinforcing other regulatory mechanisms, explaining the difficulty to phenotype strains devoid of dihydrouridines.

D-sites; how to confirm them?

Our data support that eukaryotic mRNAs are obviously dihydrouridylated. However, the 143 detected D-sites remain candidates and should be biochemically validated.

A general method to confirm the presence of a modification on RNA polymerase II transcripts consists of the isolation of polyA$^+$ RNAs followed by LC/MS. However, a polyA-based purification is tricky because it does not allow to get rid of contaminating RNAs that are abundant such as rRNAs and tRNAs (Legrand et al., 2017). Concluding from such experiments is therefore difficult and may be questionable.

Site-specific cleavage and radioactive-labeling followed by ligation-assisted extraction and thin layer chromatography (SCARLET) is a complex name for a technique with a simple goal; detecting the presence of a modified residue at the single nucleotide-resolution (**Fig. S18**). It is a very powerful quantitative approach that was originally used for m^{6}A detection but could virtually be applied to all modifications (Liu

[4] Within the framework of this study, we successfully created the catalytic inactive Dus4$_{C108A-K149A}$ and Dus2$_{C116A-K159A}$ mutants (data not shown).

and Pan, 2015). D-sites could thus be confirmed by such an approach. Moreover, this technique would provide a stoichiometric insight about the mRNA degree of modification. In D-seq, the only quantitative element is the spike-in control RNA that reveals RT termination downstream of the dihydrouridylated position for one fifth of the reads (D-ratio of ~20%), although the molecule is fully modified (100% are D, Fig. 13A). Although the D-ratio can quantify the relative stoichiometry of a dihydrouridylated position (Fig. 13E), comparing the D-ratios of spike-in RNA with mRNA D-ratios provides a limited information (Fig. 17E).

In order to confirm the detected D-sites, another conceivable experiment would be to benefit from the differential Watson-Crick interaction intensities between A-U and A-D (Table 2). As already done with tRNAs, actual dihydrouridylation of mRNAs could be assessed in a microarray-based experiment (see Introduction III.B.ii., (Peng et al., 2003; Xing et al., 2004)). In this case, the detection does not rely on a chemical labeling of the dihydrouridine and therefore would consist of a Rho-independent detection of D.

In parallel, other high-throughput techniques to detect D could be set up and comparison of the isolated candidates would confirm their relevance. These techniques should involve the precipitation of dihydrouridylated mRNAs; (i) by using an anti-dihydrouridine Ab (not available yet) or (ii) by using an anti-rhodamine Ab. The second option main drawback is that it relies on the Rho labeling as well. The immunoprecipitation of R+ RNAs was not successful (data not shown), probably because the epitope recognized by the anti-rhodamine Ab is not accessible due to the covalent binding of Rho to the RNA chain. We therefore implemented a modified Rho labeling protocol with a commercially available biotin-rhodamine molecule (instead of the rhodamine alone), followed by a precipitation with streptavidin-beads (**Fig. 29A**). Preliminary data suggest that a dihydrouridylated tRNA is indeed precipitated in R+ (or BR+ for *biotin-rhodamine effective labeling*) but not in mock-treated samples (data not shown). The implementation of a high-throughput sequencing after (immuno)precipitation of (labeled) dihydrouridines would be highly informative (**Fig. 29B**) and could be adapted in a m^6A-LAIC-seq manner to be quantitative (see Introduction IV.B.ii.) (Molinie et al., 2016).

Dus enzymes; how do they target mRNAs?

It is not surprising to observe a lack of apparent consensus sequence for the dihydrouridylation of a mRNA (Fig. 18B) despite the high substrate specificity of Dus enzymes (Fig. 7A). First, because the D-seq did not establish what enzyme for what target (briefly mentioned in the Addendum section), and secondly because the well-known tRNA Dus substrates do not display a consensus sequence either (Panwar and Raghava, 2014). Instead, a structural motif could be recognized by Dus enzymes (*e.g.* a polyuridic sequence, Fig. 18C). Interestingly, a programmed ribosomal frameshifting (PRF) of some viruses relies on a U-rich tract called the *slippery sequence* (*e.g.* 5'-UUUUUUA-3' for HIV, (Jacks et al., 1988)), upstream of a *stimulatory sequence* (*e.g.* a pseudoknot). Moreover, PRF is also known to occur in yeast and in higher eukaryotes (reviewed in (Dinman, 2012)). It would be interesting to know whether; (i) D-sites are often found upstream of stimulatory-like sequences, which is not a simple conclusion to draw because the prediction of RNA folding can be arbitrary; (ii) D-sites are often found upstream of putative novel ORFs that would be translated after the potential frameshift caused by D.

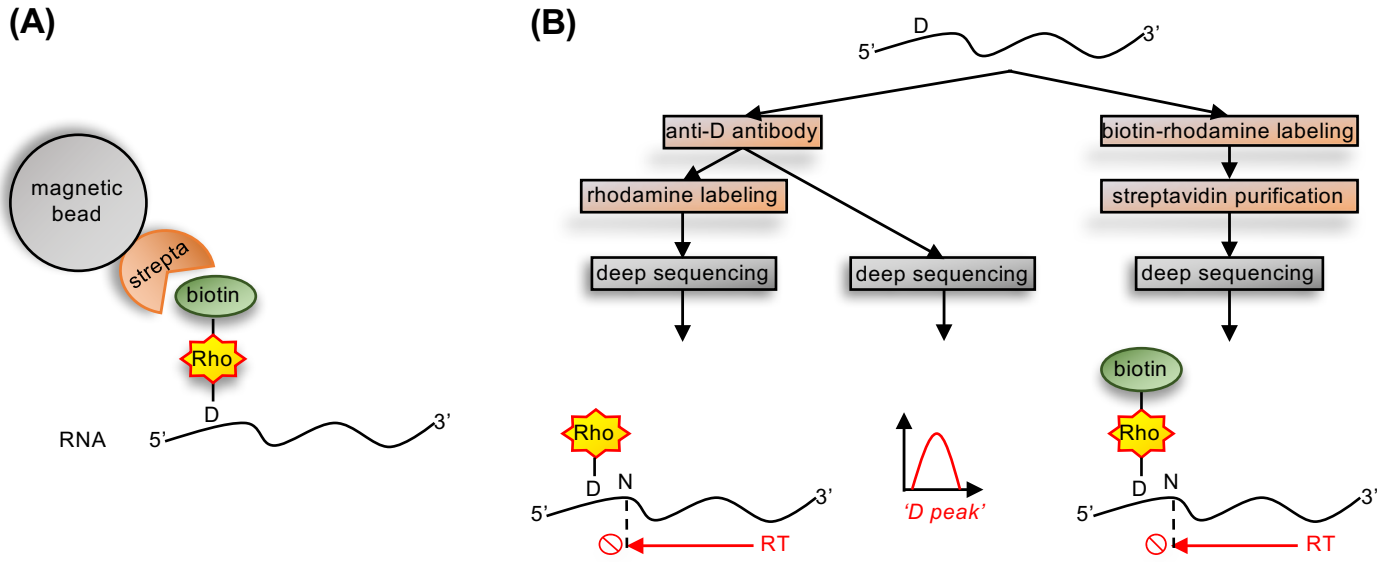


Figure 29 – Proposed pipelines to confirm D-seq candidates. (A) Precipitation of biotin-rhodamine labeled D residues is achieved by binding to streptavidin (strepta)-beads. **B)** Enriched dihydrouridylated RNAs could be sequenced after direct immunoprecipiation (anti-D antibody) or through chemical labeling (biotin-rhodamine labeling).

Bacterial Dus were shown to adopt astonishing features for substrate specificities (Fig. 8) but these conclusions cannot be extrapolated to eukaryotic Dus enzymes because hDUS2 is the sole dihydrouridine synthase with a resolved structure in Eukarya (Fig. 7E). Although the recognition mechanisms could be substantially different between tRNA and mRNA, the main question is similar in both cases; what are the molecular determinants found on the substrate for the synthesis of a dihydrouridine? Controversial pieces of evidence do not allow any absolute answer to this question. On the one hand, *in vitro* transcribed tRNAs – i.e. free of modifications – were shown to be dihydrouridylated (Byrne et al., 2015; Jiang et al., 1997; Kusuba et al., 2015; Xing et al., 2002), but on the other hand, studies suggest that a properly modified and folded tRNA is necessary for Dus catalysis (Bou-Nader et al., 2015; Cavaille et al., 1999; Yu et al., 2011). In other words, D would be a late modification in tRNA processing because the Dus enzymes require a pre-modified molecule, even though seminal works reported the presence of dihydrouridine on unspliced yeast tRNAs (Etcheverry et al., 1979; Jiang et al., 1997; Knapp et al., 1978). The truth might be that it is not a *one rule mechanism* because it was shown that dihydrouridylation occurs on *in vitro* transcript but the reaction is much faster on *in vivo* substrates, which is more or less in line with the idea that a Dus protein can bind an unmodified tRNA but cannot apparently dihydrouridylate it (Bou-Nader et al., 2015; Rider et al., 2009). But whatever the tRNA case is, it is certainly different from the mRNA dihydrouridylation. However, it can serve as a baseline for reflection; if a highly structured tRNA is indeed required for Dus recognition, it means that the D residues on mRNAs should be found on peculiar secondary structures, except if Dus enzymes are able to bind tremendously different substrates (highly structured for tRNA *vs* simply structured for mRNA) (Lewis et al., 2017). Finally, a guided dihydrouridylation process could also occur with the help of a Dus-interactome; the Dus enzymes could indeed have specific partners that influence the biology of these proteins (targets, localizations, functionality, *etc.*).

Another clue to answer to '*how do Dus enzymes target mRNAs?'* is *'when does dihydrouridylation occur?'*. Is it a co-transcriptional process? Is it nuclear, cytoplasmic, or both? To answer to this question, the localization of Dus proteins should be clearly established. Budding yeast Dus1 is likely to carry a mitochondrial motif whereas Dus3 and Dus4 are not (Xing et al., 2002). In the fission yeast however, Dus4 seems to be exclusively mitochondrial unlike the nucleocytoplasmic Dus2 and Dus3 proteins (Matsuyama et al., 2006). Once D-sites on mRNAs will be associated with one or more *dus* gene, the spatiotemporal question should be accurately investigated.

Translating a dihydrouridylated mRNA; structurally possible?

It is legitimate to ask whether it is structurally possible to translate a dihydrouridylated mRNA. Disturbing the coding ability of a transcript can result in modulation of the translational speed, impediment of translation and/or miscoding. Particularly, Watson-Crick base-pairings between positions 1 and 2 of the codon and positions 36 and 35 of the anticodon are critical for ribosomal decoding (Ogle et al., 2001). The wobble position is more *flexible* and more *tolerant* to structural mRNA modifications; it was shown that A or G methylation on positions 1 and 2 but not on position 3 of the codon impacted the translational speed and accuracy (You et al., 2017). Strikingly, almost 60% of detected D-sites were on positions 1 and 2 of codons, precisely where they can be expected if dihydrouridine is relevant for translational

regulation (**Fig. 30**). In other words, D is found at positions where it is supposed to have an impact on translation, that is to say a biological role.

The function of D could be either translation premature abortion and/or induction of ribosome stalling and/or miscoding by incorporation of non-cognate amino acids. The first hypothesis is supported by previous studies. In the late sixties, several teams investigated the *in vitro* coding properties of D-containing oligoribonucleotides. Despite the technical boundaries encountered at that time – *e.g.* the uncontrolled dihydrouridylation of the RNA template –, they concluded that the presence of D resulted in the loss of coding ability. Rottman and Cerutti showed that a ribopolymer carrying 4.2% of D (and 95.8% of U) lost up to 60% of Phe(UUU) residues incorporation into the protein – when the value for a polyU was set at 100% (Rottman and Cerutti, 1966). Another study highlighted the complete loss of ability for GUD, GDU and GDD trinucleotides to code for Val(GUU) (Smrt et al., 1966). The conclusion was the same for the dihydrouridylation of the AUG(Met) codon (Lee et al., 1967; Smrt et al., 1970). It is outstanding to see that the scientific community was already curious about the coding properties of dihydrouridine… half a century ago! They were the first to show that D-mRNA could act as a repressive translational regulator, or at least to lack the ability to be recognized as a conventional nucleotide.

To precisely dissect this hypothesis, we are currently working on an *in vitro* translation assay with an *in vitro* transcribed mRNA that contains a unique D residue at a known position (Karijolich and Yu, 2011). Our goal is to assess the stoichiometry of putative translation termination when a template is dihydrouridylated.

The missing enzyme; 'DusD' in *E. coli*?

Our data indicate that the 23S-D_{2449} deposition in *E. coli* is independent of the *dusA-B-C* genes (Fig. 20E-F). To our knowledge, this is the first experimental evidence suggesting that D could be synthesized by a *noncanonical* dihydrouridine synthase. This issue was already discussed previously and is based on Dus homologs (Bishop et al., 2002). Interestingly, the pyrimidine catabolism in bacteria and in higher eukaryotes includes the reduction of uracil to dihydrouracil by a dihydropyrimidine dehydrogenase (**Fig. 31**). In *E. coli,* the uracil reductases PreA and PreT were shown to catalyze this reaction (Hidese et al., 2011). They form a complex that could be responsible for D formation on RNA, raising the possibility of a fourth bacterial Dus enzyme. However, two major points should be addressed; (i) are these proteins able to bind RNAs? Many metabolic enzymes were recently shown to unexpectedly interact with RNAs in mammalian cells without apparent RNA-binding motifs (Castello et al., 2015). Furthermore, assays to evaluate Dus-tRNA interaction have already been set up and could be used to assess the PreA/T ability to bind (t)RNAs (Byrne et al., 2015; Yu et al., 2011), (ii) is it structurally possible for PreA/T to target nucleic acids? Their known substrate is the nucleobase alone and the RNA structure could be inaccessible for these enzymes. Crystallographic structures of dihydropyrimidine dehydrogenase and homolog enzymes are available and should, by analogy, provide a partial answer to this question (Dobritzsch et al., 2001; Rowland et al., 1998; Schnackerz et al., 2004). Noticeably, the dihydropyrimidine dehydrogenase is of clinical importance in human because deficiency in this enzyme (DPYD in *H. sapiens*) leads to a severe sensitivity to the administration of 5-fluorouracil, an agent used widely to treat cancer (Schnackerz et al., 2004).

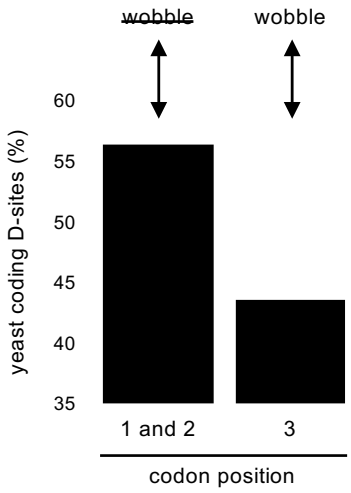
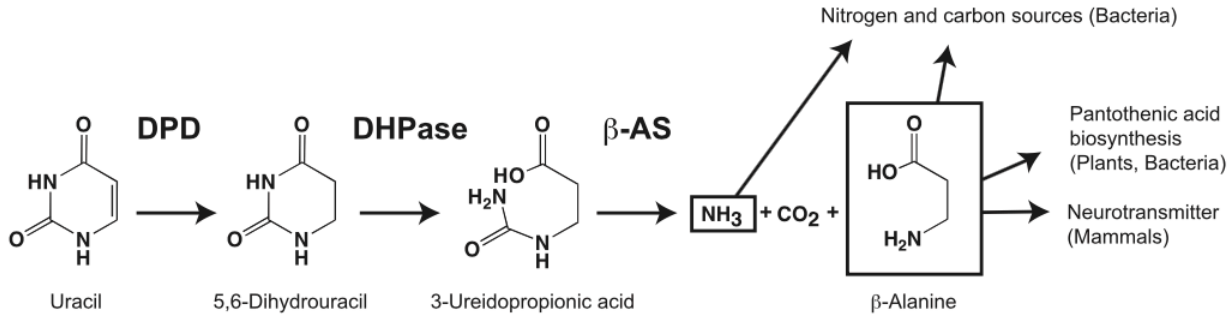


Figure 30 – D-sites are found at positions that do not face the flexible tRNA wobble position. D-seq in yeast revealed 124 D-sites on codons. 56.5% are found on positions 1 and 2 of codons, interacting with the tRNA positions 36 and 35, respectively. For information, 47.5% of uridines found on the fission yeast coding transcriptome are on position 3 of the codon.



Mammals	DPD	DHPase	β-AS
Pseudomonas aeruginosa	PydX-PydA	PydB	PydC
Escherichia coli K12	PreT-PreA	—	—

Figure 31 – Pyrimidine catabolic pathway. Metabolic degradation of a uracil nucleobase as found in bacteria and mammals. Dihydropyrimidine dehydrogenase (DPD), dihydropyrimidinase (DHPase), β-alanine synthase (β-AS) (Hidese et al., 2011).

Remarkably, the catalytic activity of this potential D writer is reversible, allowing uracil formation from dihydrouracil. In a short communication published in 1967, Kosinski and colleagues said that "[…] the presence of one or more dihydrouridylic acid residues […] together with the existence […] of a nucleotide-specific dihydrouridine dehydrogenase might constitute a translational control mechanism […]" ((Lee et al., 1967), p. 25c). Saying it with modern concepts, the dihydropyrimidine dehydrogenase activity could act as a D eraser. In agreement with the idea of reverse dihydrouridylation, an *in vitro* assay depicted that the incubation of bacterial bulk tRNA with yeast Dus2 resulted in the oxidation of dihydrouridine, in other words in a reaction where dihydrouridine is transformed into uridine (Rider et al., 2009). Dihydrouridine could therefore be a dynamic RNA modification.

Long-term perspectives; what is on the list?

The characterization of potential novel D writer or eraser has been discussed above. Another critical information to get for functional analysis of dihydrouridine would be to depict possible D-interacting proteins, the so-called *readers.* A straightforward method was used for isolation of both m^6A and m^5C readers and could be applied to find out potential D readers (Dominissini et al., 2012; Yang et al., 2017). It relies on incubation of a cell extract preparation with biotinylated RNAs containing or not the modification of interest, followed by LC-MS. Complexification of the screen – to make it more relevant – could consist of ChIRP (*chromatin isolation by RNA purification*, **Fig. S19**) experiments on WT and Δ*4* – or single *dus* mutants – strains by targeting a D-seq mRNA candidate.

The coding properties of dihydrouridine also suggest that the dihydrouridylation of a codon could lead to miscoding, that is to say to the incorporation of non-cognate amino acids. Mass spectrometric analyses of proteins generated from a D-containing RNA template would be of high interest to investigate this possibility.

Last but not least, ribosome profiling seems essential to bring insights about the ribosome behavior in the absence of a dihydrouridine (for a review see (Brar and Weissman, 2015)). Rando and colleagues recently provided ribosome occupancy maps for almost all single mutants of tRNA modifying enzymes in the budding yeast (Chou et al., 2017). Importantly, the codon occupancy was shown to be reduced for Leu codons in a Δ*dus2* mutant (**Fig. 32**), suggesting that the loss of Dus2-dependent dihydrouridylation prevents the ribosome pausing. A possible scenario is that D on mRNAs generate ribosome stalling, which is line with the nonrandom dihydrouridylation of Leu codons revealed by D-seq (Fig. 18F). It is however essential to keep in mind that this kind of technique does not allow to dissect between tRNA-dependent and -independent effects. And it is not trivial because a mutation in the D-loop can lead to miscoding, proving that the tRNA body also modulates the decoding ability (Hirsh, 1971). However, in the aforementioned study, the most affected codons in a Δ*dus2* mutant are U-containing codons, raising the possibility of an mRNA-dependent differential ribosome profiling.

mutant	site	codon	aa
Δ*dus2*	A	UUA	Leu
Δ*dus2*	A	CUG	Leu
Δ*dus2*	A	CUA	Leu
Δ*dus2*	A	CUU	Leu
Δ*dus1*	P	GCG	Ala

Figure 32 – Ribosome profiling in budding yeast. Deletion of *dus2* (mutant, column 1) results in lower codon occupancy (highlighted in red) in ribosomal A-site (column 2) for a set of codons (column 3). The cognate amino acid (aa) is depicted in column 4. Higher codon occcupancy in the mutant is highlighted in green. Only results for *dus* mutants and (-0.3 > codon occupancy > 0.3) are shown. Results from (Chou et al., 2017).

CONCLUSION

Even though the function of dihydrouridine is not well understood, its high content in tRNAs from psychrophilic bacteria or cancer cells obviously emphasizes a central role in cell adaptation.

(Bou-Nader et al., 2018)

The universal dihydrouridine RNA modification is synthesized by the highly conserved Dus enzymes. Although a precise molecular function of D is still lacking, plenty of indications were gathered by the scientific community that argue in favor of a biological relevance of this simple but fascinating posttranscriptional modification. In this study, we set up a unique transcriptome-wide mapping of dihydrouridine, permitting its detection at the single-nucleotide resolution. The efficient base modification calling on transfer and synthetic RNAs is a good evidence to trust unexpected dihydrouridylation on eukaryotic mRNAs. Remarkably, nonrandom distribution of D residues on U-rich coding regions, along with the absence of effect on the abundance of mRNAs devoid of dihydrouridine, are all in line with a concept already mentioned 50 years ago; dihydrouridine could be a regulator of translation. Thus, our unbiased investigation shed light on the importance of focusing on correlation between mRNA dihydrouridylation and translational regulation. Further experiments will help unraveling the biology of this structurally unique modification that could also be dynamically installed – and erased? – upon different physiological contexts, such as the environmental temperature. Remarkably, the D-seq pipeline could be easily applied to other RNA modifications that are sensitive to the rhodamine labeling and for which the cognate RNA modifying enzyme mutants are available. In a time when the epitranscriptomic approaches are still lacking, the pipeline proposed in this study turns out to be possibly useful for other transcriptome-wide studies. Herculean efforts have still to be made to improve the detection of RNA modifications by nanopore technologies, hence the necessity to provide epitranscriptomic tools that enable single modification detection. D-seq is virtually applicable to RNA from any organism/cell type, grown in any condition and is therefore a reliable assay for functional studies on dihydrouridine. As a proof-of concept, a bacterial transcriptome was also studied and opened the possibility of a still to be confirmed new dihydrouridine synthase. In conclusion, this epitranscriptomic study reveals the dihydrouridine as a novel eukaryotic internal mRNA modification that specifically targets coding regions.

Various experts in the field suggest that some modified nucleotides decorate mRNA as a result of the nonspecific off-target activity of enzymes that are involved in the modification of much more prevalent RNA species such as rRNA and tRNA. I believe that although this possibility cannot be disregarded, many modified nucleotides will be brought into the limelight when their cellular activities are elucidated in the coming years.

Gideon Rechavi in (Frye et al., 2016)

EXPERIMENTAL PROCEDURES

Growth conditions and spot assays.

Schizosaccharomyces pombe strains were grown in YES (Formedium PCM0310) or EMM (for Fig. S16 and Fig. 24B) (MP 4110-012) at 32°C to OD_{595nm} 0.5-0.8. For spot assays, strains were cultured in liquid YES or EMM media at 32°C and harvested at late-exponential phase, spotted (5-fold dilutions) on YES agar plates (+ chemical compound for Fig. 24) and incubated up to 5 days at 20, 25, 30, 32 or 37°C. *Escherichia coli* strains were grown at 37°C in LB (Invitrogen 12780-029) to OD_{595nm} 0.6-0.7. For spot assays, strains were cultured in liquid LB medium at 37°C and harvested at late-exponential phase, spotted (10-fold dilutions) on LB agar plates and incubated up to 2 days at 20, 30, 37 or 42°C. Human cells (HCT 116, human colorectal carcinoma, epithelial cells, ATCC CCL-247) were grown at 37°C under 5% CO_2 in McCoy's 5a medium (ATCC 30-2007) with 10% fetal bovine serum and 1% PEN-STREP.

Strains, vectors and oligo(ribo)nucleotides.

Strains are listed in **Table 4** and vectors in **Table 5**. All oligo(ribo)nucleotides were purchased from IDT and are listed in **Table 6**.

For gene conversions in *S. pombe*, Expand High Fidelity-generated PCR products (Roche 19012322) were generated by amplification of a nucleotide sequence (consisting of a resistance cassette against an antibiotic or a tag sequence cloned in a pFA6A vector) with primers designed by the Pombe PCR Primer Program (developed by the Bahler lab, http://www.bahlerlab.info/resources/, 3'-end 20nt homology with the vector and 5'-end 60nt homology with the genome). Purified PCR products (Qiagen 28106) were transformed by following the lithium acetate procedure (Murray et al., 2016). Gene deletions resulted in [ATG-STOP] replacement and gene taggings resulted in a C-terminal tagging by removal of the endogenous STOP codon and insertion of the TAP (tandem affinity purification tag) or HA (human influenza hemagglutinin tag) sequence.

dus4 and *prp4* cloning in a pREP3 vector was achieved as follows; Phusion High Fidelity-PCR amplification (NEB M0530) of the gene of interest was performed on genomic DNA, purified (Qiagen 28106), enzymatically restricted (*SalI-BamHI* for *dus4* and *SmaI-BamHI* for *prp4*), ligated (Promega M180A) into the restricted pREP3 vector, transformed in *E. coli* DH10B competent cells and sequenced. The cloning strategy resulted in expression of the gene of interest under the control of the *nmt1* promoter (*no message in thiamine 1*). Vectors were transformed on *S. pombe* cells as described above.

S. pombe GST-dus1 was prepared as follows; Phusion High Fidelity-PCR amplification (NEB M0530) of *dus1* (without its start codon) was performed on genomic DNA, purified (Qiagen 28106), enzymatically restricted (*BamHI-XhoI*), ligated (Promega M180A) into the restricted pGEX-4T1 vector (downstream and in phase with the *GST* sequence), transformed in *E. coli* DH10B competent cells and sequenced. *S. pombe GST-dus2* was prepared as follows; Phusion High Fidelity-PCR amplification

strain or cell line	#	genotype	references
Schizosaccharomyces pombe			
WT	94	*h⁻*	(Leupold, 1950)
WT*	13	*h⁻ leu1-32*	Stratagene SpQ01
WT**	1943	*h⁻ Z::nat^R-ADH15-mCherry-atb2*	(Lynch et al., 2014)
Δ*dus1*	1779	*h⁻ dus1::nat^R*	this study
Δ*dus2*	1654	*h⁻ dus2::gen^R*	this study
Δ*dus3*	1659	*h⁻ dus3::hph^R*	this study
Δ*dus4*	1493	*h⁻ dus4::ble^R*	this study
Δ*dus1-3-4*	1795	*h^{90} dus1::nat^R dus3::hph^R dus4::ble^R*	this study
Δ*du2-3-4*	1794	*h^{90} dus2::gen^R dus3::hph^R dus4::ble^R*	this study
Δ*4*	1755	*h⁻ dus1::nat^R dus2::gen^R dus3::hph^R dus4::ble^R*	this study
Δ*4**	1922	*h⁻ dus1::nat^R dus2::gen^R dus3::hph^R dus4::ble^R leu1-32*	this study
cct4-TAP	1863	*h⁻ cct4-TAP-gen^R*	this study
cct4-TAP Δ*4*	1887	*h⁻ cct4-TAP-gen^R dus1::nat^R dus2::gen^R dus3::hph^R dus4::ble^R*	this study
cct5-TAP	1864	*h⁻ cct5-TAP-gen^R*	this study
cct5-TAP Δ*4*	1888	*h⁻ cct5-TAP-gen^R dus1::nat^R dus2::gen^R dus3::hph^R dus4::ble^R*	this study
slp1-HA	1895	*h⁻ slp1-HA-gen^R*	this study
slp1-HA Δ*4*	1891	*h⁻ slp1-HA-kan^R dus1::nat^R dus2::gen^R dus3::hph^R dus4::ble^R*	this study
Δ*dus1* Δ*dus2* Δ*dus3 dus4$_{C108A-K149A}$*	/	*h⁻ dus1::nat^R dus2::gen^R dus3::hph^R dus4$_{C108A-K149A}$ Z::nat^R-ADH15-mCherry-atb2*	this study
Escherichia coli			
WT	JC8679	recombination strain	(Zhang et al, 1998)
Δ*3*	PS5107	*dusA::cm^R dusB-fis::kan^R dusC::spec^R*	(Bishop et al., 2002)

Table 4 →

strain or cell line	#	genotype	references
Homo sapiens			
WT	/	HCT 116 p53 +/+	ATCC CCL-247
DUS1 -/- (clone 1)	/	allele 1 : Δ82,064,820_82,065,063 allele 2 : Δ82,064,820_82,065,063	this study
DUS2 -/- (clone 1)	/	allele 1 : Δ68,038,060_68,038,144 allele 2 : Δ68,038,060_68,038,144	this study
DUS3 -/- (clone 1)	/	allele 1 : Δ5,791,033_5,791,152 allele 2 : Δ5,791,021_5,791,136	this study
DUS4 -/- (clone 1)	/	allele 1 : Δ107,567,102_107,567,272 allele 2 : Δ107,567,102_107,567,272	this study
HsΔ4 (clone 1)	/	*DUS*1 allele 1 : Δ82,064,820_82,065,063 *DUS*1 allele 2 : Δ82,064,022_82,064,815 *DUS*2 allele 1 : Δ68,038,060_68,038,144 *DUS*2 allele 2 : Δ68,038,060_68,038,144 *DUS*3 allele 1 : Δ5,791,037_5,791,138 *DUS*3 allele 2 : Δ5,791,037_5,791,139 *DUS*4 allele 1 : Δ107,567,102_107,567,273 *DUS*4 allele 2 : Δ107,567,102_107,567,274	this study
HsΔ4 (clone 2)	/	*DUS*1 allele 1 : Δ82,064,820_82,065,063 *DUS*1 allele 2 : Δ82,064,022_82,064,815 *DUS*2 allele 1 : Δ68,038,060_68,038,144 *DUS*2 allele 2 : Δ68,038,060_68,038,144 *DUS*3 allele 1 : Δ5,791,027_5,791,041 *DUS*3 allele 2 : Δ5,791,027_5,791,041 *DUS*4 allele 1 : Δ107,567,102_107,567,272 *DUS*4 allele 2 : Δ107,567,102_107,567,272	this study

→ **Table 4 – List of strains used in this study.** nat^R (resistance gene for noursethricin), gen^R (resistance gene for geniticin), hph^R (resistance gene for hygromycin), ble^R (resistance gene for phleomycin), cm^R (resistance gene for chloramphenicol), kan^R (resistance gene for kanamycin), $spec^R$ (resistance gene for spectinomycin), * (used for Fig. S16 and Fig. 24B only), ** (used for Fig. 25 only). The *S. pombe* strains are associated with a number (#) that represents the classification in our collection of strains. The *E. coli Δ3* mutant was generated from the JC8679 recombination strain, as explained in (Bishop et al., 2002). The human cell lines used in this study are all HCT 116 p53 +/+ (as depicted for the WT line). The CRISPR-Cas9 mediated deletion of DNA sequences (Δ) are precisely indicated with genomic coordinates for each targeted allele (HCT 116 cells are diploid).

usage	#	features
as depicted in Fig. S16 and Fig. 24B		
empty	pDH79	pREP3-*pnmt1*
dus4	pDH788	pREP3-*pnmt1-dus4*
prp4	pDH789	pREP3-*pnmt1-prp4*
for protein purification		
GST-dus1	pDH809	pGEX-4T1-*GST-dus1*
GST-dus2	pDH810	pGEX-4T1-*GST-dus2*$_{cDNA}$

Table 5 – List of vectors used in this study. pDH (plasmid of Damien Hermand lab), *GST* (sequence coding for glutathione-S-transferase), *pnmt1* (promoter repressed by the presence of thiamine).

usage	#	5'-sequence-3'	F/R
Schizosaccharomyces pombe			
resistance cassette gene	244	CGGATGTGATGTGAGAACTGTATCCTAGC	R
dus1 deletion check	2049	GCTACAAATCTACAAAATCCTGCC	F
dus2 deletion check	1950	GAGAATCGTTAATTGTGTCCG	F
dus3 deletion check	2061	TGTGTACCTAAAAGCATATAGCG	F
dus4 deletion check	2065	TAGTAGAAACGCCAACTGG	F
$tRNA_i^{Met}$ primer extension	2540	ACCTACGGGTTATGAGCCCGTCGG	R
$tRNA^{AspGUC}$ (*SPCTRNAASP.06*) primer extension	2743	GGGCTGCAAGCGTGAC	R
SPMITTRNAARG.02 primer extension	2746	CTAACAATTTCGAAGATTGTTAC	R
$tRNA^{PheGAA}$ northern blotting	2552	TGACCAACAGATCTTCAGTCTGTC	R
$tRNA^{CysGCA}$ northern blotting	3026	AGGGACCAGCCGGATTTGAACCA	R
5S rRNA northern blotting	2202	TTCCCATGTTGTCTCCAACC	R
cct4-TAP for Expand PCR	2837	CAGCTGGCTGCTGAAACGACCAAGATGATCATGAAGATTGACGACATTACCTTAGCTCGT-CGGATCCCCGGGTTAATTAA	F
cct4-TAP for Expand PCR	2838	ACAATGTAAATTCTTCTTTTCTCTTTCCTGTAATCCAATTTTCGGAAGGGAGACGAAACC-GAATTCGAGCTCGTTTAAAC	R
cct4-TAP tagging check	2850	GCGGGTATTAATGTTCGTAAGGG	F
cct5-TAP for Expand PCR	2839	AGAATGGTACTCAAAGTTAACGATATTATCGTTGCCGGTTCTAAGGATGATGACTATAAC-CGGATCCCCGGGTTAATTAA	F
cct5-TAP for Expand PCR	2840	GCTTGCCGTATGAAATGAAGTAAATTAAATTTCCTTTAGCCGTGCGGATGAAAATGCCAT-GAATTCGAGCTCGTTTAAAC	R
cct5-TAP tagging check	2851	GCAGACCGGAAGTAATGATATGAG	F
slp1-HA for Expand PCR	2873	GACCACGTTAAAAGGCCCATTCCAATTACCAAAACCCCGTCCAGCAGCATAACAATCCGT-CGGATCCCCGGGTTAATTAA	F
slp1-HA for Expand PCR	2874	CTGATGCAACATAGTGATTCGACGAGCTAAACGTTGAAATCTAAAAGAAACTGGTGTTGT-GAATTCGAGCTCGTTTAAAC	R
slp1-HA tagging check	2875	GCAGCCAGTGATGAGAATCTG	F
nda2 qRT-PCR	1242	CTGTTTCATCCTGAACAGATGG	F
nda2 qRT-PCR	1243	GCCATACGACGTATCCTTTC	R
cct1 qRT-PCR	2824	CCACAAAGGGCATTGATGATC	F
cct1 qRT-PCR	2825	CTTGGCGATACGTCTCAAATC	R
act1 qRT-PCR	739	CCACTATGTATCCCGGTATTGC	F
act1 qRT-PCR	740	CAATCTTGACCTTCATGGAGCT	R
25S rRNA qRT-PCR	815	TTGTCCATGAAATTCCATTGAA	F
25S rRNA qRT-PCR	816	CTTACAATACCCCGTTCCACAT	R
nab3 qRT-PCR	2363	GAAAGTATAATGTCAAACTGGTATCCA	F
nab3 qRT-PCR	2364	CTATCTGTGCTAAAGGCCCATAA	R
adh1 qRT-PCR	785	GAAGGAAGCCGACATGATTG	F
adh1 qRT-PCR	786	CAGCTTGCTCGTAAGACTTGG	R
dus4 cloning in pREP3	2667	GGCAGCGTCGACATGAGAGATCGTTTAAAGGACCCTG	R
dus4 cloning in pREP3	2668	ACACGGGGATCCTTAATAGCGGCGACAAGGC	F
prp4 cloning in pREP3	2716	ACACGGGGATCCATGAGTGACGATAGATTTGC	F
prp4 cloning in pREP3	2717	TCCCCCGGGGGATTATTTTTTTATAAAGAAAGGATGCTTC	R
dus1 cloning for GST-Dus1 production	2932	CGCGGATCCGCGGCTTCAAAAAAGCTTCATGG	F
dus1 cloning for GST-Dus1 production	2933	CCGCTCGAGCGGCTATGCCACAGAAGCATTGAC	R
gBLOCK $dus2_{cDNA}$	/	https://www.pombase.org/gene/SPBC1709.06, exons sequence	F
dus2 cloning for GST-Dus2 production	2934	CGCGGATCCGCGGGTCTGCTAAATTATAGCAATAAAG	F
dus2 cloning for GST-Dus2 production	2935	CCGCTCGAGCGGCTAAACGACTTGAGCTTTCTCTTC	R
Escherichia coli			
23S rRNA primer extension	3080	ATATGAACTCTTGGGCGGTATC	R

Table 6 →

usage	#	5'-sequence-3'	F/R
Homo sapiens			
DUS1 deletion check	/	AATTGAGCTGTCTCTGGAAAGG	F
DUS1 deletion check	/	CTGCCTCTGAGTTGAGTCTTAAG	R
DUS2 deletion check	/	GGCTGTAACAGAGGAGGAAATG	F
DUS2 deletion check	/	CAGTGTCACACATGTTTGTCTC	R
DUS3 deletion check	/	ATCTCAGGTACCTCCGCTGAC	F
DUS3 deletion check	/	GAAGGTCTTCCCGCCACAC	R
DUS4 deletion check	/	ATATATGTTGTTGGCAGTTGAAATTG	F
DUS4 deletion check	/	CATATATACTAGGCAAGTAGTAGCATC	R
DUS1 crRNA	/	CCCTCATCGTGCAGGTGCGCAGG	F
DUS1 crRNA	/	GCAGCTTTGGCATCGTCTCCAGG	R
DUS2 crRNA	/	GGGCCAGGATTAGCTTATTATGG	F
DUS2 crRNA	/	AGCGGACATTGTTTACTGTGAGG	R
DUS3 crRNA	/	TAAGCGTCAGTGAGTCGCCGGGG	F
DUS3 crRNA	/	TCCGCCGTTCCCTCCGCCATCGG	R
DUS4 crRNA	/	CATGCAAACGACAATATGTCAGG	F
DUS4 crRNA	/	GTTTGAAGTTCCCATATGGTGGG	R
$tRNA^{PheGAA}$ primer extension	2963	AGATCTTCAGTCTAACGCTC	R
$tRNA^{GlyGCC}$ primer extension	2960	GAATCGAACCCGGGCCTCCC	R
$tRNA_i^{MetAUG}$ primer extension	2962	TGGTAGCAGAGGATGGTTTCGATC	R
$tRNA^{HisGUG}$ primer extension	2964	GAGGTTGCTGCGGCCACAAC	R
Other			
gBLOCK for synthetic RNA	/	**TAATACGACTCACTATAG**GGGAGGCGAGAACACACCACAACGAAAACGAGCAAAACCCGG*TA*CGCAACACAAAAGC GAACAACGCGAAAAAGGACACCGAAGCGGAAGCAAAGACAACCAACAGAAAACAACCGCAAACAAACGGGACCAGA CAACGCACCAGCAAAA **T7P**, first transcribed nucleotide, *unique transcribed U (or D)*	/
gBLOCK for synthetic RNA with 3 D	/	**TAATACGACTCACTATAG**GGGAGGCGAGAACACACCACAACGAAAACGAGCAAAACCC*TTT*ACGCACCACAAAAGC GAACAACGCGAAAAAGGACACCGAAGCGGAAGCAAAGACAACCAACAGAAAACAACCGCAAACAAACGGGACCAGA CAACGCACCAGCAAAA **T7P**, first transcribed nucleotide, *transcribed U (or D)*	
gBLOCK PCR	2576	TAATACGACTCACTATAGGGG	F
gBLOCK PCR	2577	TTTTGCTGGTGCGTTG	R
synthetic RNA primer extension	2567	GTCTTTGCTTCCGCTTCGGTGTCC	R
synthetic RNA with 3 D primer extension	3079	CGCGTTGTTCGCTTTTGTGG	R
D-seq			
RNA adapter	2661	5'-P-AGATCGGAAGAGCGTCGTG-ddC	/
DNA primer for RT	2662	ACACGACGCTCTTCCGA	R
DNA adapter	2674	5'-P-AGATCGGAAGAGCACACGTCTG-ddC	/
universal PCR primer	2679	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	F
barcoded PCR primers	/	CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	R

→ Table 6 – List of DNA primers, RNA primers and synthetic gene fragments used in this study. RNA oligonucleotides are shown in blue, ddC (dideoxycytidine), 5'-P (5'-phosphate extremity), F (forward), R (reverse), T7P (promoter for T7 RNA polymerase). The barcoded primer is characterized by a 6nt-long index (XXXXXX in red), which corresponds to the reversed sequence provided by the TruSeq Stranded Total RNA Sample Preparation Guide (Illumina). For example, the barcoded PCR primer AR005 (Illumina ACAGTG) will contain the CACTGT sequence instead of XXXXXX.

(NEB M0530) of *dus2* cDNA sequence (without its start codon) was performed on a gBLOCK gene fragment (IDT), purified (Qiagen 28106), enzymatically restricted (*BamHI-XhoI*), ligated (Promega M180A) into the restricted pGEX-4T1 vector (downstream and in phase with the *GST* sequence), transformed in *E. coli* DH10B competent cells and sequenced. Vectors were eventually transformed in *E. coli* BL21 cells for protein expression in LB + 1mM IPTG at 37°C for 120min and recombinant proteins were purified by combining the Poly-Prep Chromatography Columns (Bio-Rad 731-1550) with Glutathione Sepharose 4B (GE 17-0756-01).

Gene inactivation by implementation of CRISPR-Cas9 in human cells.

CRISPR-Cas9 mediated deletion was performed as described by IDT (Alt-R CRISPR-Cas9 System) but with important modifications (Dr. F. Ernst, Pr. Lafontaine lab, ULB). The strategy relies on the electroporation of ribonucleoprotein complexes (RNPs). First, 100µM of RNA duplexes were formed by mixing equal volumes of 200µM crRNA and 200µM of tracrRNA for 5min at 95°C, then cooled for 15min at RT. Each electroporation reaction required a 25µL-RNP solution and 1.75 x 10^6 of HCT 116 cells. The RNP solution was prepared by mixing the following for 20min at RT; 5µL of 1X PBS, 10µL of Cas9 enzyme (612 pmol) and 10µL of the crRNA:tracrRNA duplexes (for single-gene targeting: 5µL of crRNA forward:tracrRNA + 5µL of crRNA reverse:tracrRNA, for quadruple-gene targeting: 4 x 1.25µL of crRNA forward:tracrRNA + 4 x 1.25µL of crRNA reverse:tracrRNA). Subcultivation ratio of 1:10 was implemented 2-3 days before the CRISPR experiment. 1.75 x 10^6 of early passaged HCT 116 cells were centrifuged for 5min at 120*g*, washed in 1X PBS, resuspended in 100µL of Nucleofector solution (Lonza VCA-1003) and added to the RNP-solution. The RNP/cells mix was transferred to a Nucleofector cuvette with 5µL of Electroporation Enhancer, electroporated on Nucleofector Device (D-032 program) and immediately supplemented with 400µL of RPMI medium. After 10min of incubation at RT, the electroporation product was divided into three replicative wells and incubated for 24h at 37°C under 5% CO_2 with 825µL of McCoy's 5a + FBS medium. Primary diagnostic consisted of PCR amplification(s) of the targeted locus/loci on the so-called *CRISPR population* to assess the generation of a DNA fragment that was shorter than the WT band (revealing expected deletion). Individual clones were obtained by diluting the CRISPR population on a 96-well plate. The clones were then diagnosed with PCRs and the homozygous candidates were sequenced and checked for *Mycoplasma* contamination (GATC Biotech).

RNA extraction.

50 mL of *S. pombe* (OD_{595nm} ~ 0.5) or *E. coli* (OD_{595nm} ~ 0.6) were harvested by centrifugation and washed in DEPC water. Humans cells (confluency ~ 90%) were collected by trypsinization and washed in 1X PBS. The cell pellets were resuspended in 750µL of TES buffer (10mM Tris pH 7.5, 10mM EDTA pH 8, 0.5% SDS) and the RNA was isolated by hot phenol extraction; equal volume of phenol:chloroform 5:1 was added (Sigma P1944), heated and shaken at 65°C for 60min, the upper phase was collected after centrifugation, mixed with an equal volume of phenol:chloroform:IAA 125:24:1 (Sigma 77619); the upper phase was then collected after centrifugation, mixed with an equal volume of chloroform:IAA 24:1 (Sigma 25666); the upper phase was collected again and the RNA was alcohol-precipitated with 2.5 volumes of 100% ethanol and 0.1 volume of 3M sodium acetate pH 5.2 for minimum 30min at -80°C or

overnight at -20°C. After centrifugation, the pellet was washed with 70% ethanol, air-dried and resuspended in DEPC (Sigma D5758) water or pure water (Sigma 900682).

Microchip analysis of RNA.

Total, chemically treated and ribodepleted RNA samples were visualized by RNA 6000 Pico Assay by using the Bioanalyzer microchip technology according to the manufacturer's instructions (Agilent 5067-1513). Total RNAs were expected to have a RIN (RNA integrity number) > 8.5 and > 6 for Rho labeled RNAs.

Rhodamine labeling and dot blot assay.

Rhodamine labeling was performed as previously described with slight modifications (Betteridge et al., 2007). Briefly, 30µg of total RNA were resuspended in a 400µL volume with 0.04M Tris-HCl pH 7.4 and 4mg of sodium borohydride (Sigma 452882, freshly resuspended in 1mM potassium hydroxide) for 60min at 25°C in dark with constant shaking (750rpm). Mock samples (R-) were treated similarly but without sodium borohydride. The reaction was stopped by addition of glacial acetic acid to a final concentration of ~0.3M. The reduced RNA was then ethanol-precipitated as described above and the dried pellet was resuspended in 5µL of DEPC water and 85µL of 0.1M sodium formate pH 4 (Sigma 71539) by heating at 65°C for 7min. Fluorescent labeling was performed for 90min at 37°C in dark with 2.2mM of rhodamine 110 chloride (Sigma 83695) or 2.2mM of biotin-rhodamine 110 (Biotium 80022). 10X dye concentrates were prepared in advance in 100% methanol and conserved at -20°C in dark for maximum 26 months. The pH of the reaction was then lowered to 7.5 with 0.8M Tris-HCl pH 8.5 and the dye excess was eliminated by phenol extraction (Sigma P1944) for 60min at 65°C. After centrifugation, the labeled RNA-containing upper phase was collected and RNA was ethanol-precipitated as described above. For dot blot assay, RNA was resuspended in < 10µL of DEPC water by heating at 65°C and 0.2 to 1µg/µL of RNA was spotted on a Hybond-N^+ membrane (GE RPN1210b). The signal intensity was immediately monitored with a Cy3 channel (GE ImageQuant LAS 4000) and the quantification was performed with ImageJ (Schneider et al., 2012). The intensity of a DEPC water drop was considered as the background, subtracted to all other values and the subsequent signal of a 22mM rhodamine 110 in methanol drop was scaled to 1. Methylene blue staining was used as a loading control; the membrane was incubated for 5min at RT with 0.5M sodium acetate and 0.04% methylene blue, then washed twice in 1X PBS-T.

***in vitro* dihydrouridylation.**

S. cerevisiae GST-Dus1 was a gift of E.M. Phizicky (URMC, NY, US). *in vitro* dihydrouridylation was based on a previously published protocol (Xing et al., 2002). 30µg of total RNA were incubated for 40min at 30°C with 1µg GST-Dus, 100mM Tris-HCl pH 8, 100mM ammonium acetate, 5mM magnesium chloride, 2mM DTT, 0.1mM EDTA, 1mM NADH and 1mM NADPH. The mock reaction (depicted as '+ *buffer*') was done with 10mM glutathione instead of GST-Dus (GE 27-4570-01). The dihydrouridylated RNA was then purified with an equal volume of phenol:chloroform:IAA 125:24:1 (Sigma 77619), ethanol-precipitated as described above and resuspended in DEPC water for subsequent rhodamine labeling.

Synthetic RNA preparation.

Synthetic RNA was prepared similarly to (Schwartz et al., 2014). An ordered gBLOCK gene fragment (IDT, Table 6) was resuspended in 1X TE buffer. 20ng were amplified with Phusion High-Fidelity DNA Polymerase (NEB M0530S), the subsequent PCR product was selected on an agarose gel and purified with QIAquick PCR Purification Kit (Qiagen 28106). For RNA production, 200ng of DNA were *in vitro* transcribed by following the RiboMAX™ Large Scale RNA Production Systems' instructions (Promega P1300). For D spike-in, rUTP was replaced by rDTP (Tebu-bio 040N-1035-1). The synthetic RNA was purified with ProbeQuant™ G-50 Micro Columns by following the manufacturer's instructions (GE 28-9034-08). A total of 30ng of spike-in was added to 30µg of total RNA for primer extension and 3ng for D-seq.

Primer extension.

Primer extension was performed essentially as described (Sakurai et al., 2010). First, 10pmol of 5'-unphosphorylated primer were ^{32}P-labeled with 10U of PNK (Promega M4101) and 3µL of [ɣ-^{32}P]-ATP (Perkin NEG002A250UC). The reaction was performed at 37°C for 10min and stopped at 90°C for 2min. The resulting labeled primer was diluted to 0.4pmol/µL. 5 to 25µg of denatured R+/R- RNA were retro-transcribed for 60min at 50°C with 150µM dNTPs (Applied biosystems 4368814), 1µL of ^{32}P-labeled primer and 100U of SSIII in 1X first-strand buffer (Invitrogen 18080-044). The reaction was stopped by addition of 2X sample buffer (98% formamide, 10mM EDTA, 0.1% blue bromophenol, 0.1% xylene cyanol) and incubation for 10min at 90°C. The samples were immediately resolved on a 15 to 20% PAGE with 7M urea at 225V in 1X TBE buffer (Sigma T4415) and visualized with Super RX-N film (Fujifilm 47410 19289). The ladder was composed of DNA oligonucleotides (0.5 µM each) that were ^{32}P-labeled as described above and mixed with 2X sample buffer. For Fig. 23A, the aniline-dependent RT termination at D sites was performed as described previously (Marchand et al., 2018).

Fluorescent Northern blotting.

10µg of R+/R- or untreated total RNA were denatured in 1mM Tris-HCl pH 7.4, 6mM EDTA, 6% glycerol and were electrophoresed on a 1.2% agarose gel (0.67% formaldehyde, 20mM MOPS, 5mM sodium acetate, 1mM EDTA) in 1X FA buffer (20mM MOPS, 5mM sodium acetate, 1mM EDTA, pH 7). The gel was washed once in 50mM sodium hydroxide and twice in neutralization buffer (1.5M sodium chloride, 0.5M Tris-HCl pH 7.4). The nucleic acids were capillary transferred overnight in 1X SSC (Sigma S6639) on a Hybond-XL membrane (GE RPN203S). After crosslinking for 120min at 80°C, the rhodamine signal was monitored as described above. Methylene blue staining was used as a loading control; the membrane was incubated for 5min at RT with 0.5M sodium acetate and 0.04% methylene blue, then washed twice in 1X PBS-T.

tRNA Northern blotting.

Small RNA fraction was enriched from 200-400µg of total RNA by incubation 30min on ice with 190mM sodium chloride and 7.3% PEG 8000. The supernatant was carefully collected after centrifugation for 20min at 15,500*g* at 4°C and ethanol-precipitated as described above. The pellet was washed with 95% ethanol, air-dried

and resuspended in DEPC water. 5µg of small RNA were denatured in 1X sample buffer (50% formamide, 2.5mM EDTA, 0.01% blue bromophenol), electrophoresed on a 10% PAGE with 7M urea at 225V in 1X TBE buffer (Sigma T4415) and transferred on a Hybond-XL membrane (GE RPN203S) by semi-dry transfer for 45min at 3mA/cm^2 (Thermo Owl HEP Series Semidry Electroblotting System). Membrane was rinsed in 2X SSC before UV crosslinking for 30sec at 254nm (Vilbert Lourmat). For probe preparation, 20pmol of 5'-unphosphorylated primer were ^{32}P-labeled with 10U of PNK (Promega M4101) and 2µL of [ɣ-^{32}P]-ATP (Perkin NEG002A250UC) in a 20µL-reaction for 60 min at 37°C. The probe was purified using ProbeQuant™ G-50 Micro Columns following the manufacturer's instructions (GE 28-9034-08), denatured and hybridized to the crosslinked membrane overnight at 42°C in PerfectHyb™ Plus Hybridization Buffer (Sigma H7033). The membrane was then washed in 2X SSC 0.1% SDS, twice in 0.1X SSC 0.1% SDS at 42°C and visualized with Super RX-N film (Fujifilm 47410 19289).

D-seq library preparation.

Libraries for sequencing were prepared by combining previously published pipelines (Engreitz et al., 2013; Shishkin et al., 2015). 2-5µg of R+/R- total RNA were ribodepleted (Thermo A15020 and K155005 [modified protocol provided by the manufacturer] for *S. pombe* and *H. sapiens* and Illumina MRZGN126 for *E. coli*), quantified with Qubit RNA HS assay (Thermo Q32852) and analyzed by microchip as described above. Zinc-mediated fragmentation was performed at 70°C for 15min with 90-125 ng of ribodepleted RNA and the reaction was stopped by following the manufacturer's instructions (Thermo AM8740). Simultaneous DNA digestion and RNA extremities dephosphorylation were performed at 37°C for 30min with 2U of TURBO DNase (Thermo AM2238) and 3U of FastAP Thermosensitive Alkaline Phosphatase (Thermo EF0654) in 10mM Tris-HCl pH 7.4, 1mM magnesium chloride, 0.12mM calcium chloride, 10mM potassium chloride, 0.002% Triton X-100 and 2mM DTT. 40U of RNase inhibitor were added to prevent RNA degradation (NEB M0314). 20pmol of RNA adapter were then added to the sample, followed by denaturation for 2min at 70°C. 3'-end ligation was performed for 90min at 25°C with 39U of T4 RNA ligase (NEB M0437) and 12U of RNase inhibitor (NEB M0314) in 1X NEB Ligase 1 Buffer, 9% DMSO, 1mM ATP and 20% PEG 8000. After ligation, 10pmol of DNA primer were added, the sample was denatured for 2min at 70°C and the reverse transcription was performed as described above with SSIII (Invitrogen 18080-044). The excess of DNA primer was subsequently digested with 2µL of ExoSAP-IT (Affymetrix 78250) by incubating 4min at 37°C and the enzyme was then inactivated by incubating 1min at 80°C. The RNA was degraded for 10min at 70°C with 100mM sodium hydroxide and the solution was neutralized by addition of acetic acid. 40pmol of DNA adapter were then added to the sample, followed by denaturation for 2min at 75°C. The 3'-end ligation was performed overnight at 25°C with 48U of T4 RNA ligase (NEB M0437) in 1X NEB Ligase 1 Buffer, 4% DMSO, 1mM ATP and 23.75% PEG 8000. Nucleic acid clean-ups were performed between each step by using MyOne SILANE magnetic beads (Thermo 37002D) resuspended in the chaotropic Buffer RLT (Qiagen 74104). PCR amplification was performed in a 50µL-reaction by mixing cDNA, 1µM of barcoded PCR primer, 1µM of universal PCR primer and 25µL of NEBNext High-Fidelity 2X PCR Master Mix (NEB M0541) with the following program; 30sec at 98°C, 5 cycles of (10sec at 98°C, 30sec at 67°C, 30sec at 72°C), 13 cycles of (10sec at 98°C, 30sec at 72°C) and 60sec at 72°C as a final extension. Finally, two successive rounds of DNA polymer size selection were performed by mixing PCR products with 1.2 and

1.1 volume of AMPure XP beads (Beckman A63880). The DNA library was quantified (Thermo Q32851), examined on a DNA microchip (Agilent 5067-4626, ~250-300nt polymers) and sequenced on Illumina HiSeq 2500 [1 flow cell, 2 lanes, 300M clusters] or on Novaseq S1 [100 cycles, 1 flow cell, 2 lanes, 1600M clusters] (paired-end reads, 2 x 50nt).

Gene list and motif analyses.

The *S. pombe* lists of genes were analyzed by the fission yeast-specific platform called AnGeLi (Bitton et al., 2015). The sequence motif analyses were performed with the web-based application WebLogo (Crooks et al., 2004).

Codon occurrence in *S. pombe* transcriptome.

The probability to get a D-site on the coding *S. pombe* transcriptome was calculated as follows; there are 27 codons containing a single U, representing 1,099,338 occurrences, 9 codons containing two U residues, representing 612,597 occurrences and 1 codon containing three U residues, representing 92,872 occurrences. There are therefore 2,603,148 possibilities to get a D-site on a coding region [(1 x 1,099,338) + (2 x 612,597) + (3 x 92,872) = 2,603,148]. Expressed in percentages, single U, double U and triple U-containing codons represent 42.2, 47.1 and 10.7% of the possibilities to get a D-site, respectively.

qRT-PCR.

0.5µg of RNeasy-purified total RNA (Qiagen 74104) was retro-transcribed with the High-Capacity cDNA Reverse Transcription Kit (Thermo 4368813) by following the manufacturer's instructions. The real-time PCR amplification was performed with SYBR Green Supermix (Bio-Rad 172-5124) in a Bio-Rad CFX96™ Real-Time machine. The PCR program was 3min at 95°C and 40 cycles of (15sec at 95°C and 30sec at 60°C). Relative RNA quantification relied on the $\Delta\Delta C_T$ method.

Western blotting.

Proteins were extracted from 10mL of yeasts cultured in YES at 32°C. Cells were pelleted, washed in water and disrupted with 0.3M hydroxide sodium for 10min at RT. The lysate was then centrifuged and the supernatant was discarded. The pellet was resuspended in 70µL of alkaline extraction buffer (60mM Tris-HCl pH 6.8, 4% β-mercaptoethanol, 4% SDS, 0.01% bromophenol blue, 5% glycerol) and boiled for 5min at 95°C. 10µL of the samples were loaded on a 4-15% Mini-PROTEAN TGX Precast Protein Gel (Bio-Rad 456-1083), transferred on a nitrocellulose membrane (Bio-Rad 1704158) and blocked for 60min at RT or overnight at 4°C in 50:50 1X PBS:Odyssey Blocking Buffer PBS (Westburg LI 927-40100). The membrane was incubated with the primary antibody for 60min at RT or overnight at 4°C with Odyssey Buffer containing 0.05% Tween20 (Bio-Rad 161-0781) and a 1:500-1,000 dilution of anti-tubulin (Sigma T5168), anti-HA (Sigma H6908) or anti-TAP (Sigma P1291) primary antibodies. After three washes in PBS-T, the membrane was incubated with a 1:10,000 dilution of anti-mouse (Westburg LI 925-32210) or anti-rabbit (Westburg LI 925-32211) secondary antibodies for 60min at RT in Odyssey Buffer containing 0.05% Tween20. After three washes in PBS-T and three washes in PBS, the membrane was dried for 60min at

37°C and visualized on a LiCOR scanner on channel 800. The quantification was performed with ImageJ (Schneider et al., 2012).

RNA-sequencing.

1µg of total RNA was ribodepleted as described in the D-seq section and analyzed by microchip electrophoresis as described above. 125ng of ribodepleted samples were prepared for strand-specific NGS with the TruSeq Stranded Total RNA Sample Preparation Protocol (Illumina). The library was sequenced on Illumina HiSeq 2500 (paired-end reads, 2 x 50nt). Reads were mapped on the genome using TopHat2. The quantification was made with featureCounts function on R and the differential expression analysis was performed with DESeq2. The genes following these criteria were considered as differentially expressed; false discovery rate < 10% and absolute value of the fold change > 1.5.

BIBLIOGRAPHY

I. FIGURES- AND TABLES-ASSOCIATED REFERENCES

Figures	References
1	(Megel et al., 2015)
2	(Li et al., 2018; Quinn and Chang, 2016)
3	(Boccaletto et al., 2018; Chawla et al., 2015; Grosjean, 2009; Jackman and Alfonzo, 2013)
4	(Holley et al., 1965; Moreau et al., 2011; Saenger, 1984; Whelan et al., 2015)
7	(Bou-Nader et al., 2018; Bou-Nader et al., 2015; Byrne et al., 2015; Chen et al., 2013; Griffiths et al., 2012; Kasprzak et al., 2012; Park et al., 2004; Whelan et al., 2015; Yu et al., 2011)
9	(Dominissini et al., 2012; Grozhik and Jaffrey, 2018; Helm and Motorin, 2017; Meyer et al., 2012; Song and Yi, 2017)
31	(Hidese et al., 2011)
32	(Chou et al., 2017)

Tables	References
1	(Dalluge et al., 1996b; Davis et al., 1986; Deb et al., 2014; Deslauriers et al., 1971; Formoso and Tinoco, 1971; Jack et al., 1976; Quigley and Rich, 1976; Rohrer and Sundaralingam, 1970; Stuart et al., 1996; Suck et al., 1971, 1972; Sundaralingam et al., 1971a, b; Westhof et al., 1985; Westhof and Sundaralingam, 1986; Woo et al., 1980)
2	(Batt et al., 1954; Beltchev and Grunberg-Manago, 1970; Betteridge et al., 2007; Cerutti et al., 1968a; Cerutti et al., 1968b; Cerutti and Miller, 1967; Cohn and Doherty, 1955; Jacobson and Hedgcoth, 1970; Kaur et al., 2011; Magrath and Shaw, 1967; Molinaro et al., 1968; Pan et al., 2009; Peng et al., 2003; Wintermeyer and Zachau, 1971; Xing et al., 2004)
3	(Beltchev and Grunberg-Manago, 1970; Cerutti et al., 1968a; Cerutti and Miller, 1967; Igo-Kemenes and Zachau, 1969; Macon and Wolfenden, 1968; Molinaro et al., 1968; Schleich et al., 1978; Wintermeyer et al., 1979; Wintermeyer and Zachau, 1970, 1971, 1974, 1979)
4	(Bishop et al., 2002; Leupold, 1950; Lynch et al., 2014; Zhang et al., 1998)

Supplemental	References
S1	(Boccaletto et al., 2018)
S3	(Vare et al., 2017)
S4	(Krog et al., 2011)
S18	(Liu and Pan, 2015)
S19	(Chu et al., 2011)

II. LIST OF REFERENCES

Agarwala, S.D., Blitzblau, H.G., Hochwagen, A., and Fink, G.R. (2012). RNA methylation by the MIS complex regulates a cell fate decision in yeast. PLoS Genet *8*, e1002732.

Alarcon, C.R., Lee, H., Goodarzi, H., Halberg, N., and Tavazoie, S.F. (2015). N6-methyladenosine marks primary microRNAs for processing. Nature *519*, 482-485.

Alexandrov, A., Chernyakov, I., Gu, W., Hiley, S.L., Hughes, T.R., Grayhack, E.J., and Phizicky, E.M. (2006). Rapid tRNA decay can result from lack of nonessential modifications. Mol Cell *21*, 87-96.

Alexandrov, A., Martzen, M.R., and Phizicky, E.M. (2002). Two proteins that form a complex are required for 7-methylguanosine modification of yeast tRNA. RNA *8*, 1253-1266.

Arcari, P., and Brownlee, G.G. (1980). The nucleotide sequence of a small (3S) seryl-tRNA (anticodon GCU) from beef heart mitochondria. Nucleic Acids Res *8*, 5207-5212.

Bar-Yaacov, D., Mordret, E., Towers, R., Biniashvili, T., Soyris, C., Schwartz, S., Dahan, O., and Pilpel, Y. (2017). RNA editing in bacteria recodes multiple proteins and regulates an evolutionarily conserved toxin-antitoxin system. Genome Res *27*, 1696-1703.

Basanta-Sanchez, M., Temple, S., Ansari, S.A., D'Amico, A., and Agris, P.F. (2016). Attomole quantification and global profile of RNA modifications: Epitranscriptome of human neural stem cells. Nucleic Acids Res *44*, e26.

Batt, R.D., J.K., M., J.M., P., and J., M. (1954). Chemistry of dihydropyrimidines. Ultraviolet spectra and alkaline decomposition. J Am Chem Soc *76*, 3663-3665.

Bauer, F., Matsuyama, A., Candiracci, J., Dieu, M., Scheliga, J., Wolf, D.A., Yoshida, M., and Hermand, D. (2012). Translational control of cell division by Elongator. Cell Rep *1*, 424-433.

Becker, H.F., Motorin, Y., Florentz, C., Giege, R., and Grosjean, H. (1998). Pseudouridine and ribothymidine formation in the tRNA-like domain of turnip yellow mosaic virus RNA. Nucleic Acids Res *26*, 3991-3997.

Behm-Ansmant, I., Helm, M., and Motorin, Y. (2011). Use of specific chemical reagents for detection of modified nucleotides in RNA. J Nucleic Acids *2011*, 408053.

Beltchev, B., and Grunberg-Manago, M. (1970). Preparation of a pG-fragment from tRNA(Phe)(yeast) by chemical scission at the dihydrouracil, and inhibition of tRNA(Phe)(yeast) charging by this fragment when combined with the -CCA half of this tRNA. FEBS Lett *12*, 24-26.

Best, A.N. (1978). Composition and Characterization of tRNA from Methanococcus vannielii. J Bacteriol *133*, 240-250.

Betteridge, T., Liu, H., Gamper, H., Kirillov, S., Cooperman, B.S., and Hou, Y.M. (2007). Fluorescent labeling of tRNAs for dynamics experiments. RNA *13*, 1594-1601.

Birkedal, U., Christensen-Dalsgaard, M., Krogh, N., Sabarinathan, R., Gorodkin, J., and Nielsen, H. (2015). Profiling of ribose methylations in RNA by high-throughput sequencing. Angew Chem Int Ed Engl *54*, 451-455.

Bishop, A.C., Xu, J., Johnson, R.C., Schimmel, P., and de Crecy-Lagard, V. (2002). Identification of the tRNA-dihydrouridine synthase family. J Biol Chem *277*, 25090-25095.

Bitton, D.A., Schubert, F., Dey, S., Okoniewski, M., Smith, G.C., Khadayate, S., Pancaldi, V., Wood, V., and Bahler, J. (2015). AnGeLi: A Tool for the Analysis of Gene Lists from Fission Yeast. Front Genet *6*, 330.

Boccaletto, P., Machnicka, M.A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T.K., de Crecy-Lagard, V., Ross, R., Limbach, P.A., Kotter, A.*, et al.* (2018). MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Res *46*, D303-D307.

Borek, E., Baliga, B.S., Gehrke, C.W., Kuo, C.W., Belman, S., Troll, W., and Waalkes, T.P. (1977). High turnover rate of transfer RNA in tumor tissue. Cancer Res *37*, 3362-3366.

Bou-Nader, C., Montemont, H., Guerineau, V., Jean-Jean, O., Bregeon, D., and Hamdane, D. (2018). Unveiling structural and functional divergences of bacterial tRNA dihydrouridine synthases: perspectives on the evolution scenario. Nucleic Acids Res *46*, 1386-1394.

Bou-Nader, C., Pecqueur, L., Bregeon, D., Kamah, A., Guerineau, V., Golinelli-Pimpaneau, B., Guimaraes, B.G., Fontecave, M., and Hamdane, D. (2015). An extended dsRBD is required for post-transcriptional modification in human tRNAs. Nucleic Acids Res *43*, 9446-9456.

Bourgon, R., Gentleman, R., and Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. Proc Natl Acad Sci U S A *107*, 9546-9551.

Brar, G.A., and Weissman, J.S. (2015). Ribosome profiling reveals the what, when, where and how of protein synthesis. Nat Rev Mol Cell Biol *16*, 651-664.

Bregeon, D., Colot, V., Radman, M., and Taddei, F. (2001). Translational misreading: a tRNA modification counteracts a +2 ribosomal frameshift. Genes Dev *15*, 2295-2306.

Buchser, W.J., Smith, R.P., Pardinas, J.R., Haddox, C.L., Hutson, T., Moon, L., Hoffman, S.R., Bixby, J.L., and Lemmon, V.P. (2012). Peripheral nervous system genes expressed in central neurons induce growth on inhibitory substrates. PLoS One *7*, e38101.

Buck, M., Connick, M., and Ames, B.N. (1983). Complete analysis of tRNA-modified nucleosides by high-performance liquid chromatography: the 29 modified nucleosides of Salmonella typhimurium and Escherichia coli tRNA. Anal Biochem *129*, 1-13.

Byrne, R.T., Jenkins, H.T., Peters, D.T., Whelan, F., Stowell, J., Aziz, N., Kasatsky, P., Rodnina, M.V., Koonin, E.V., Konevega, A.L.*, et al.* (2015). Major reorientation of tRNA substrates defines specificity of dihydrouridine synthases. Proc Natl Acad Sci U S A *112*, 6033-6037.

Cahova, H., Winz, M.L., Hofer, K., Nubel, G., and Jaschke, A. (2015). NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. Nature *519*, 374-377.

Calvo, S.E., Clauser, K.R., and Mootha, V.K. (2016). MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res *44*, D1251-1257.

Cantara, W.A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendeix, F.A., Fabris, D., and Agris, P.F. (2011). The RNA Modification Database, RNAMDB: 2011 update. Nucleic Acids Res *39*, D195-201.

Carlile, T.M., Rojas-Duran, M.F., Zinshteyn, B., Shin, H., Bartoli, K.M., and Gilbert, W.V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature *515*, 143-146.

Carr, D.O., and Grisolia, S. (1964). Incorporation of Dihydrouridine Monophosphate and Uridine Monophosphate into Liver and Brain Ribonucleic Acid. J Biol Chem *239*, 160-166.

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C.A., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A.*, et al.* (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res *44*, D471-480.

Castello, A., Hentze, M.W., and Preiss, T. (2015). Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. Trends Endocrinol Metab *26*, 746-757.

Cavaille, J., Chetouani, F., and Bachellerie, J.P. (1999). The yeast Saccharomyces cerevisiae YDL112w ORF encodes the putative 2'-O-ribose methyltransferase catalyzing the formation of Gm18 in tRNAs. RNA *5*, 66-81.

Cech, T.R., and Steitz, J.A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. Cell *157*, 77-94.

Cerutti, P., Holt, J.W., and Miller, N. (1968a). Detection and determination of 5,6-dihydrouridine and 4-thiouridine in transfer ribonucleic acid from different sources. J Mol Biol *34*, 505-518.

Cerutti, P., Kondo, Y., Landis, W.R., and Witkop, B. (1968b). Photoreduction of uridine and reduction of dihydrouridine with sodium borohydride. J Am Chem Soc *90*, 771-775.

Cerutti, P., Miles, H.T., and Frazier, J. (1966). Interaction of partially reduced polyuridylic acid with polyadenylic acid. Biochem Biophys Res Commun *22*, 466-472.

Cerutti, P., and Miller, N. (1967). Selective reduction of yeast transfer ribonucleic acid with sodium borohydride. J Mol Biol *26*, 55-66.

Chan, C.T., Deng, W., Li, F., DeMott, M.S., Babu, I.R., Begley, T.J., and Dedon, P.C. (2015). Highly Predictive Reprogramming of tRNA Modifications Is Linked to Selective Expression of Codon-Biased Genes. Chem Res Toxicol *28*, 978-988.

Chan, C.T., Dyavaiah, M., DeMott, M.S., Taghizadeh, K., Dedon, P.C., and Begley, T.J. (2010). A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. PLoS Genet *6*, e1001247.

Chawla, M., Oliva, R., Bujnicki, J.M., and Cavallo, L. (2015). An atlas of RNA base pairs involving modified nucleobases with optimal geometries and accurate energies. Nucleic Acids Res *43*, 6714-6729.

Chen, K., Lu, Z., Wang, X., Fu, Y., Luo, G.Z., Liu, N., Han, D., Dominissini, D., Dai, Q., Pan, T.*, et al.* (2015). High-resolution N(6) -methyladenosine (m(6) A) map using photo-crosslinking-assisted m(6) A sequencing. Angew Chem Int Ed Engl *54*, 1587-1590.

Chen, M., Yu, J., Tanaka, Y., Tanaka, M., Tanaka, I., and Yao, M. (2013). Structure of dihydrouridine synthase C (DusC) from Escherichia coli. Acta Crystallogr Sect F Struct Biol Cryst Commun *69*, 834-838.

Chen, P., Jager, G., and Zheng, B. (2010). Transfer RNA modifications and genes for modifying enzymes in Arabidopsis thaliana. BMC Plant Biol *10*, 201.

Chen, X., Sim, S., Wurtmann, E.J., Feke, A., and Wolin, S.L. (2014). Bacterial noncoding Y RNAs are widespread and mimic tRNAs. RNA *20*, 1715-1724.

Chen, Y.G., Kowtoniuk, W.E., Agarwal, I., Shen, Y., and Liu, D.R. (2009). LC/MS analysis of cellular RNA reveals NAD-linked RNA. Nat Chem Biol *5*, 879-881.

Chernyakov, I., Whipple, J.M., Kotelawala, L., Grayhack, E.J., and Phizicky, E.M. (2008). Degradation of several hypomodified mature tRNA species in Saccharomyces cerevisiae is mediated by Met22 and the 5'-3' exonucleases Rat1 and Xrn1. Genes Dev *22*, 1369-1380.

Chi, K.R. (2017). The RNA code comes into focus. Nature *542*, 503-506.

Cho, B.K., Knight, E.M., Barrett, C.L., and Palsson, B.O. (2008). Genome-wide analysis of Fis binding in Escherichia coli indicates a causative role for A-/AT-tracts. Genome Res *18*, 900-910.

Chou, H.J., Donnard, E., Gustafsson, H.T., Garber, M., and Rando, O.J. (2017). Transcriptome-wide Analysis of Roles for tRNA Modifications in Translational Regulation. Mol Cell *68*, 978-992 e974.

Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell *44*, 667-678.

Cohn, W.E., and Doherty, D.G. (1955). The catalytic hydrogenation of pyrimidine nucleosides and nucleotides and the isolation of their ribose and respective ribose phosphates. J Am Chem Soc *78*, 2863-2866.

Couvillion, M.T., Bounova, G., Purdom, E., Speed, T.P., and Collins, K. (2012). A Tetrahymena Piwi bound to mature tRNA 3' fragments activates the exonuclease Xrn2 for RNA processing in the nucleus. Mol Cell *48*, 509-520.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561-563.

Crick, F.H. (1958). On protein synthesis. Symp Soc Exp Biol *12*, 138-163.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Dai, Q., Moshitch-Moshkovitz, S., Han, D., Kol, N., Amariglio, N., Rechavi, G., Dominissini, D., and He, C. (2017). Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. Nat Methods *14*, 695-698.

Dalluge, J.J., Hamamoto, T., Horikoshi, K., Morita, R.Y., Stetter, K.O., and McCloskey, J.A. (1997). Posttranscriptional modification of tRNA in psychrophilic bacteria. J Bacteriol *179*, 1918-1923.

Dalluge, J.J., Hashizume, T., and McCloskey, J.A. (1996a). Quantitative measurement of dihydrouridine in RNA using isotope dilution liquid chromatography-mass spectrometry (LC/MS). Nucleic Acids Res *24*, 3242-3245.

Dalluge, J.J., Hashizume, T., Sopchik, A.E., McCloskey, J.A., and Davis, D.R. (1996b). Conformational flexibility in RNA: the role of dihydrouridine. Nucleic Acids Res *24*, 1073-1079.

Davis, D.R. (1998). Biophysical and conformational properties of modified nucleotides in RNA. Modifiction and Editing of RNA. American Society for Microbiology Press, 85-102.

Davis, D.R., Griffey, R.H., Yamaizumi, Z., Nishimura, S., and Poulter, C.D. (1986). 15N-labeled tRNA. Identification of dihydrouridine in Escherichia coli tRNAfMet, tRNALys, and tRNAPhe by 1H-15N two-dimensional NMR. J Biol Chem *261*, 3584-3587.

Davis, F.F., and Allen, F.W. (1957). Ribonucleic acids from yeast which contain a fifth nucleotide. J Biol Chem *227*, 907-915.

de Bruijn, M.H., and Klug, A. (1983). A model for the tertiary structure of mammalian mitochondrial transfer RNAs lacking the entire 'dihydrouridine' loop and stem. EMBO J *2*, 1309-1321.

de Bruijn, M.H., Schreier, P.H., Eperon, I.C., Barrell, B.G., Chen, E.Y., Armstrong, P.W., Wong, J.F., and Roe, B.A. (1980). A mammalian mitochondrial serine transfer RNA lacking the "dihydrouridine" loop and stem. Nucleic Acids Res *8*, 5213-5222.

de Smit, M.H., Gultyaev, A.P., Hilge, M., Bink, H.H., Barends, S., Kraal, B., and Pleij, C.W. (2002). Structural variation and functional importance of a D-loop-T-loop interaction in valine-accepting tRNA-like structures of plant viral RNAs. Nucleic Acids Res *30*, 4232-4240.

Deb, I., Sarzynska, J., Nilsson, L., and Lahiri, A. (2014). Rapid communication capturing the destabilizing effect of dihydrouridine through molecular simulations. Biopolymers *101*, 985-991.

Delatte, B., Wang, F., Ngoc, L.V., Collignon, E., Bonvin, E., Deplus, R., Calonne, E., Hassabi, B., Putmans, P., Awe, S.*, et al.* (2016). RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science *351*, 282-285.

Deslauriers, R., Lapper, R.D., and Smith, I.C. (1971). A proton magnetic resonance study of the molecular conformation of a modified nucleoside from transfer RNA. Dihydrouridine. Can J Biochem *49*, 1279-1284.

Desrosiers, R., Friderici, K., and Rottman, F. (1974). Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. Proc Natl Acad Sci U S A *71*, 3971-3975.

Dinman, J.D. (2012). Mechanisms and implications of programmed translational frameshifting. Wiley Interdiscip Rev RNA *3*, 661-673.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F.*, et al.* (2012). Landscape of transcription in human cells. Nature *489*, 101-108.

Dobritzsch, D., Schneider, G., Schnackerz, K.D., and Lindqvist, Y. (2001). Crystal structure of dihydropyrimidine dehydrogenase, a major determinant of the pharmacokinetics of the anti-cancer drug 5-fluorouracil. EMBO J *20*, 650-660.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M.*, et al.* (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201-206.

Dominissini, D., Nachtergaele, S., Moshitch-Moshkovitz, S., Peer, E., Kol, N., Ben-Haim, M.S., Dai, Q., Di Segni, A., Salmon-Divon, M., Clark, W.C.*, et al.* (2016). The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. Nature *530*, 441-446.

Dudin, O., Merlini, L., Bendezu, F.O., Groux, R., Vincenzetti, V., and Martin, S.G. (2017). A systematic screen for morphological abnormalities during fission yeast sexual reproduction identifies a mechanism of actin aster formation for cell fusion. PLoS Genet *13*, e1006721.

Dudley, E., and Bond, L. (2014). Mass spectrometry analysis of nucleosides and nucleotides. Mass Spectrom Rev *33*, 302-331.

Dyubankova, N., Sochacka, E., Kraszewska, K., Nawrot, B., Herdewijn, P., and Lescrinier, E. (2015). Contribution of dihydrouridine in folding of the D-arm in tRNA. Org Biomol Chem *13*, 4960-4966.

Edelheit, S., Schwartz, S., Mumbach, M.R., Wurtzel, O., and Sorek, R. (2013). Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. PLoS Genet *9*, e1003602.

Edmonds, C.G., Crain, P.F., Gupta, R., Hashizume, T., Hocart, C.H., Kowalak, J.A., Pomerantz, S.C., Stetter, K.O., and McCloskey, J.A. (1991). Posttranscriptional modification of tRNA in thermophilic archaea (Archaebacteria). J Bacteriol *173*, 3138-3148.

Edmonds, C.G., Vestal, M.L., and McCloskey, J.A. (1985). Thermospray liquid chromatography-mass spectrometry of nucleosides and of enzymatic hydrolysates of nucleic acids. Nucleic Acids Res *13*, 8197-8206.

Ehrlich, A., Funk, C., and Merritt, A.J. (1952). The isolation of hydrouracil from beef spleen. Arch Biochem Biophys *35*, 468-469.

El Yacoubi, B., Bailly, M., and de Crecy-Lagard, V. (2012). Biosynthesis and function of posttranscriptional modifications of transfer RNAs. Annu Rev Genet *46*, 69-95.

Emmerechts, G., Barbe, S., Herdewijn, P., Anne, J., and Rozenski, J. (2007). Post-transcriptional modification mapping in the Clostridium acetobutylicum 16S rRNA by mass spectrometry and reverse transcriptase assays. Nucleic Acids Res *35*, 3494-3503.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S.*, et al.* (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973.

Etcheverry, T., Colby, D., and Guthrie, C. (1979). A precursor to a minor species of yeast tRNASer contains an intervening sequence. Cell *18*, 11-26.

Fauquenoy, S., Migeot, V., Finet, O., Yague-Sanz, C., Khorosjutina, O., Ekwall, K., and Hermand, D. (2018). Repression of Cell Differentiation by a cis-Acting lincRNA in Fission Yeast. Curr Biol *28*, 383-391 e383.

Feederle, R., and Schepers, A. (2017). Antibodies specific for nucleic acid modifications. RNA Biol *14*, 1089-1098.

Felden, B., Hanawa, K., Atkins, J.F., Himeno, H., Muto, A., Gesteland, R.F., McCloskey, J.A., and Crain, P.F. (1998). Presence and location of modified nucleotides in Escherichia coli tmRNA: structural mimicry with tRNA acceptor branches. EMBO J *17*, 3188-3196.

Findeiss, S., Langenberger, D., Stadler, P.F., and Hoffmann, S. (2011). Traces of post-transcriptional RNA modifications in deep sequencing data. Biol Chem *392*, 305-313.

Formoso, C., and Tinoco, I., Jr. (1971). Minor nucleosides in RNA: Optical studies of dinucleoside phosphates containing dihydrouridine. Biopolymers *10*, 1533-1541.

Frye, M., Jaffrey, S.R., Pan, T., Rechavi, G., and Suzuki, T. (2016). RNA modifications: what have we learned and where are we headed? Nat Rev Genet *17*, 365-372.

Gehrke, C.W., and Kuo, K.C. (1990). Ribonucleoside analysis by reversed-phase high performance liquid chromatography. Journal of chromatography library *45A*, A3-A64.

Geisler, S., and Coller, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol *14*, 699-712.

Gilbert, W. (1986). Origin of life: the RNA world. Nature *319*, 618.

Grasteau, A., Tremblay, Y.D., Labrie, J., and Jacques, M. (2011). Novel genes associated with biofilm formation of Actinobacillus pleuropneumoniae. Vet Microbiol *153*, 134-143.

Green, M., and Cohen, S.S. (1957). Studies on the biosynthesis of bacterial and viral pyrimidines. II. Dihydrouracil and dihydrothymine nucleosides. J Biol Chem *225*, 397-407.

Griffiths, S., Byrne, R.T., Antson, A.A., and Whelan, F. (2012). Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of human dihydrouridine synthase. Acta Crystallogr Sect F Struct Biol Cryst Commun *68*, 333-336.

Grosjean, H. (2009). Nucleic acids are not boring long polymers of only four types of nucleotides: a guided tour. DNA and RNA modification enzymes: structure, mechanism, function and evolution. Landes Bioscience, 1-18.

Grosjean, H., Houssier, C., Romby, P., and Marquet, R. (1998). Modulation role of modified nucleotides in RNA loop-loop interaction. American Society for Microbiology Press, 113-134.

Grozhik, A.V., and Jaffrey, S.R. (2018). Distinguishing RNA modifications from noise in epitranscriptome maps. Nat Chem Biol *14*, 215-225.

Gu, J., Patton, J.R., Shimba, S., and Reddy, R. (1996). Localization of modified nucleotides in Schizosaccharomyces pombe spliceosomal small nuclear RNAs: modified nucleotides are clustered in functionally important regions. RNA *2*, 909-918.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell *35*, 849-857.

Gupta, R., and Woese, C.R. (1980). Unsual modification patterns in the transfer ribonucleic acids of Archaebacteria. Current Microbiology *4*, 245-249.

Harvey, R., Dezi, V., Pizzinga, M., and Willis, A.E. (2017). Post-transcriptional control of gene expression following stress: the role of RNA-binding proteins. Biochem Soc Trans *45*, 1007-1014.

He, C. (2010). Grand challenge commentary: RNA epigenetics? Nat Chem Biol *6*, 863-865.

Helm, M., and Motorin, Y. (2017). Detecting RNA modifications in the epitranscriptome: predict and validate. Nat Rev Genet *18*, 275-291.

Hey, A., Li, M.S., Hudson, M.J., Langford, P.R., and Kroll, J.S. (2013). Transcriptional profiling of Neisseria meningitidis interacting with human epithelial cells in a long-term in vitro colonization model. Infect Immun *81*, 4149-4159.

Hidese, R., Mihara, H., Kurihara, T., and Esaki, N. (2011). Escherichia coli dihydropyrimidine dehydrogenase is a novel NAD-dependent heterotetramer essential for the production of 5,6-dihydrouracil. J Bacteriol *193*, 989-993.

Hirsh, D. (1971). Tryptophan transfer RNA as the UGA suppressor. J Mol Biol *58*, 439-458.

Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R., and Zamir, A. (1965). Structure of a Ribonucleic Acid. Science *147*, 1462-1465.

House, C.H., and Miller, S.L. (1996). Hydrolysis of dihydrouridine and related compounds. Biochemistry *35*, 315-320.

Huang, R.C., and Bonner, J. (1965). Histone-bound RNA, a component of native nucleohistone. Proc Natl Acad Sci U S A *54*, 960-967.

Igo-Kemenes, T., and Zachau, H.G. (1969). On the specificity of the reduction of transfer ribonucleic acids with sodium borohydride. Eur J Biochem *10*, 549-556.

Jack, A., Ladner, J.E., and Klug, A. (1976). Crystallographic refinement of yeast phenylalanine transfer RNA at 2-5A resolution. J Mol Biol *108*, 619-649.

Jackman, J.E., and Alfonzo, J.D. (2013). Transfer RNA modifications: nature's combinatorial chemistry playground. Wiley Interdiscip Rev RNA *4*, 35-48.

Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J., and Varmus, H.E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature *331*, 280-283.

Jacobson, M., and Hedgcoth, C. (1970). Determination of 5,6-dihydrouridine in ribonucleic acid. Anal Biochem *34*, 459-469.

Jaffrey, S.R. (2014). An expanding universe of mRNA modifications. Nat Struct Mol Biol *21*, 945-946.

Jiang, H.P., Chu, J.M., Lan, M.D., Liu, P., Yang, N., Zheng, F., Yuan, B.F., and Feng, Y.Q. (2016). Comprehensive profiling of ribonucleosides modification by affinity zirconium oxide-silica composite monolithic column online solid-phase microextraction - Mass spectrometry analysis. J Chromatogr A *1462*, 90-99.

Jiang, H.Q., Motorin, Y., Jin, Y.X., and Grosjean, H. (1997). Pleiotropic effects of intron removal on base modification pattern of yeast tRNAPhe: an in vitro study. Nucleic Acids Res *25*, 2694-2701.

Johnson, J.D., and Horowitz, J. (1971). Characterization of ribosomes and RNAs from Mycoplasma hominis. Biochim Biophys Acta *247*, 262-279.

Jones, C.I., Spencer, A.C., Hsu, J.L., Spremulli, L.L., Martinis, S.A., DeRider, M., and Agris, P.F. (2006). A counterintuitive Mg2+-dependent and modification-assisted functional folding of mitochondrial tRNAs. J Mol Biol *362*, 771-786.

Kadaba, S., Krueger, A., Trice, T., Krecic, A.M., Hinnebusch, A.G., and Anderson, J. (2004). Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae. Genes Dev *18*, 1227-1240.

Kadaba, S., Wang, X., and Anderson, J.T. (2006). Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. RNA *12*, 508-521.

Kadumuri, R.V., and Janga, S.C. (2018). Epitranscriptomic Code and Its Alterations in Human Disease. Trends Mol Med.

Kamenski, P., Kolesnikova, O., Jubenot, V., Entelis, N., Krasheninnikov, I.A., Martin, R.P., and Tarassov, I. (2007). Evidence for an adaptation mechanism of mitochondrial translation via tRNA import from the cytosol. Mol Cell *26*, 625-637.

Karijolich, J., and Yu, Y.T. (2011). Converting nonsense codons into sense codons by targeted pseudouridylation. Nature *474*, 395-398.

Kasprzak, J.M., Czerwoniec, A., and Bujnicki, J.M. (2012). Molecular evolution of dihydrouridine synthases. BMC Bioinformatics *13*, 153.

Kato, T., Daigo, Y., Hayama, S., Ishikawa, N., Yamabuki, T., Ito, T., Miyamoto, M., Kondo, S., and Nakamura, Y. (2005). A novel human tRNA-dihydrouridine synthase involved in pulmonary carcinogenesis. Cancer Res *65*, 5638-5646.

Kaur, J., Raj, M., and Cooperman, B.S. (2011). Fluorescent labeling of tRNA dihydrouridine residues: Mechanism and distribution. RNA *17*, 1393-1400.

Keith, G., Heitzler, J., el Adlouni, C., Glasser, A.L., Fix, C., Desgres, J., and Dirheimer, G. (1993). The primary structure of cytoplasmic initiator tRNA(Met) from Schizosaccharomyces pombe. Nucleic Acids Res *21*, 2949.

Kennedy, E.M., Bogerd, H.P., Kornepati, A.V., Kang, D., Ghoshal, D., Marshall, J.B., Poling, B.C., Tsai, K., Gokhale, N.S., Horner, S.M.*, et al.* (2016). Posttranscriptional m(6)A Editing of HIV-1 mRNAs Enhances Viral Gene Expression. Cell Host Microbe *19*, 675-685.

Kim, N.K., Theimer, C.A., Mitchell, J.R., Collins, K., and Feigon, J. (2010). Effect of pseudouridylation on the structure and activity of the catalytically essential P6.1 hairpin in human telomerase RNA. Nucleic Acids Res *38*, 6746-6756.

Kirpekar, F., Hansen, L.H., Mundus, J., Tryggedsson, S., Teixeira Dos Santos, P., Ntokou, E., and Vester, B. (2018). Mapping of ribosomal 23S ribosomal RNA modifications in Clostridium sporogenes. RNA Biol *15*, 1060-1070.

Knapp, G., Beckmann, J.S., Johnson, P.F., Fuhrman, S.A., and Abelson, J. (1978). Transcription and processing of intervening sequences in yeast tRNA genes. Cell *14*, 221-236.

Kohanski, M.A., Dwyer, D.J., Hayete, B., Lawrence, C.A., and Collins, J.J. (2007). A common mechanism of cellular death induced by bactericidal antibiotics. Cell *130*, 797-810.

Kowalak, J.A., Bruenger, E., and McCloskey, J.A. (1995). Posttranscriptional modification of the central loop of domain V in Escherichia coli 23 S ribosomal RNA. J Biol Chem *270*, 17758-17764.

Kowalak, J.A., Dalluge, J.J., McCloskey, J.A., and Stetter, K.O. (1994). The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles. Biochemistry *33*, 7869-7876.

Krog, J.S., Espanol, Y., Giessing, A.M., Dziergowska, A., Malkiewicz, A., Ribas de Pouplana, L., and Kirpekar, F. (2011). 3-(3-amino-3-carboxypropyl)-5,6-dihydrouridine is one of two novel post-transcriptional modifications in tRNALys(UUU) from Trypanosoma brucei. FEBS J *278*, 4782-4796.

Krol, A., Gallinaro, H., Lazar, E., Jacob, M., and Branlant, C. (1981). The nuclear 5S RNAs from chicken, rat and man. U5 RNAs are encoded by multiple genes. Nucleic Acids Res *9*, 769-787.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell *31*, 147-157.

Kuchino, Y., and Borek, E. (1978). Tumour-specific phenylalanine tRNA contains two supernumerary methylated bases. Nature *271*, 126-129.

Kusuba, H., Yoshida, T., Iwasaki, E., Awai, T., Kazayama, A., Hirata, A., Tomikawa, C., Yamagami, R., and Hori, H. (2015). In vitro dihydrouridine formation by tRNA dihydrouridine synthase from Thermus thermophilus, an extreme-thermophilic eubacterium. J Biochem *158*, 513-521.

Lafontaine, D., Vandenhaute, J., and Tollervey, D. (1995). The 18S rRNA dimethylase Dim1p is required for pre-ribosomal RNA processing in yeast. Genes Dev *9*, 2470-2481.

Lee, S., Brown, G.L., and Kosinski, Z. (1967). Loss of coding properties of the oligonucleotide adenylyluridylylguanosine after photoreduction or hydration. Biochem J *103*, 25C-27C.

Legrand, C., Tuorto, F., Hartmann, M., Liebers, R., Jacob, D., Helm, M., and Lyko, F. (2017). Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs. Genome Res *27*, 1589-1596.

Leupold, U. (1950). Die Verebung von Homothallie und Heterothallie bei Schizosaccharomyces pombe. C.R. Lab. Carlsberg *24*, 381-475.

Lewis, C.J., Pan, T., and Kalsotra, A. (2017). RNA modifications and structures cooperate to guide RNA-protein interactions. Nat Rev Mol Cell Biol *18*, 202-210.

Li, J.B., Levanon, E.Y., Yoon, J.K., Aach, J., Xie, B., Leproust, E., Zhang, K., Gao, Y., and Church, G.M. (2009). Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science *324*, 1210-1213.

Li, S., Xu, Z., and Sheng, J. (2018). tRNA-Derived Small RNA: A Novel Regulatory Small Non-Coding RNA. Genes (Basel) *9*.

Li, X., Xiong, X., Wang, K., Wang, L., Shu, X., Ma, S., and Yi, C. (2016a). Transcriptome-wide mapping reveals reversible and dynamic N(1)-methyladenosine methylome. Nat Chem Biol *12*, 311-316.

Li, X., Xiong, X., and Yi, C. (2016b). Epitranscriptome sequencing technologies: decoding RNA modifications. Nat Methods *14*, 23-31.

Liang, W.D., Bi, Y.T., Wang, H.Y., Dong, S., Li, K.S., and Li, J.S. (2013). Gene expression profiling of Clostridium botulinum under heat shock stress. Biomed Res Int *2013*, 760904.

Liao, Y., Castello, A., Fischer, B., Leicht, S., Foehr, S., Frese, C.K., Ragan, C., Kurscheid, S., Pagler, E., Yang, H.*, et al.* (2016). The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. Cell Rep *16*, 1456-1469.

Linder, B., Grozhik, A.V., Olarerin-George, A.O., Meydan, C., Mason, C.E., and Jaffrey, S.R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat Methods *12*, 767-772.

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015). N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature *518*, 560-564.

Liu, N., and Pan, T. (2015). Probing RNA Modification Status at Single-Nucleotide Resolution in Total RNA. Methods Enzymol *560*, 149-159.

Lo, R.Y., Bell, J.B., and Roy, K.L. (1982). Dihydrouridine-deficient tRNAs in Saccharomyces cerevisiae. Nucleic Acids Res *10*, 889-902.

Lombard, M., and Hamdane, D. (2017). Flavin-dependent epitranscriptomic world. Arch Biochem Biophys *632*, 28-40.

Lorenz, C., Lunse, C.E., and Morl, M. (2017). tRNA Modifications: Impact on Structure and Thermal Adaptation. Biomolecules *7*.

Lovejoy, A.F., Riordan, D.P., and Brown, P.O. (2014). Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. PLoS One *9*, e110799.

Luo, X., Wang, C.Z., Chen, J., Song, W.X., Luo, J., Tang, N., He, B.C., Kang, Q., Wang, Y., Du, W.*, et al.* (2008). Characterization of gene expression regulated by American ginseng and ginsenoside Rg3 in human colorectal cancer cells. Int J Oncol *32*, 975-983.

Lynch, E.M., Groocock, L.M., Borek, W.E., and Sawin, K.E. (2014). Activation of the gamma-tubulin complex by the Mto1/2 complex. Curr Biol *24*, 896-903.

Mace, K., and Gillet, R. (2016). Origins of tmRNA: the missing link in the birth of protein synthesis? Nucleic Acids Res *44*, 8041-8051.

Machnicka, M.A., Olchowik, A., Grosjean, H., and Bujnicki, J.M. (2014). Distribution and frequencies of post-transcriptional modifications in tRNAs. RNA Biol *11*, 1619-1629.

Macon, J.B., and Wolfenden, R. (1968). 1-Methyladenosine. Dimroth rearrangement and reversible reduction. Biochemistry *7*, 3453-3458.

Madison, J.T., and Holley, R.W. (1965). The Presence of 5,6-Dihydrouridylic Acid in Yeast "Soluble" Ribonucleic Acid. Biochem Biophys Res Commun *18*, 153-157.

Magrath, D.I., and Shaw, D.C. (1967). The occurrence and source of beta-alanine in alkaline hydrolysates of sRNA: a sensitive method for the detection and assay of 5,6-dihydrouracil residues in RNA. Biochem Biophys Res Commun *26*, 32-37.

Marchand, V., Ayadi, L., Ernst, F.G.M., Hertler, J., Bourguignon-Igel, V., Galvanin, A., Kotter, A., Helm, M., Lafontaine, D.L.J., and Motorin, Y. (2018). AlkAniline-Seq: Profiling of m(7) G and m(3) C RNA Modifications at Single Nucleotide Resolution. Angew Chem Int Ed Engl.

Marchand, V., Blanloeil-Oillo, F., Helm, M., and Motorin, Y. (2016). Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. Nucleic Acids Res *44*, e135.

Matsuyama, A., Arai, R., Yashiroda, Y., Shirai, A., Kamata, A., Sekido, S., Kobayashi, Y., Hashimoto, A., Hamamoto, M., Hiraoka, Y.*, et al.* (2006). ORFeome cloning and global analysis of protein localization in the fission yeast Schizosaccharomyces pombe. Nat Biotechnol *24*, 841-847.

McCloskey, J.A. (1986). Nucleoside modification in Archaebacterial transfer rna. System. Appl. Microbiol. *7*, 246-252.

McCutchan, T., Silverman, S., Kohli, J., and Soll, D. (1978). Nucleotide sequence of phenylalanine transfer RNA from Schizosaccharomyces pombe: implications for transfer RNA recognition by yeast phenylalanyl-tRNA synthetase. Biochemistry *17*, 1622-1628.

Megel, C., Morelle, G., Lalande, S., Duchene, A.M., Small, I., and Marechal-Drouard, L. (2015). Surveillance and cleavage of eukaryotic tRNAs. Int J Mol Sci *16*, 1873-1893.

Meyer, K.D., and Jaffrey, S.R. (2014). The dynamic epitranscriptome: N6-methyladenosine and gene expression control. Nat Rev Mol Cell Biol *15*, 313-326.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell *149*, 1635-1646.

Mittelstadt, M., Frump, A., Khuu, T., Fowlkes, V., Handy, I., Patel, C.V., and Patel, R.C. (2008). Interaction of human tRNA-dihydrouridine synthase-2 with interferon-induced protein kinase PKR. Nucleic Acids Res *36*, 998-1008.

Molinaro, M., Sheiner, L.B., Neelon, F.A., and Cantoni, G.L. (1968). Effect of chemical modification of dihydrouridine in yeast transfer ribonucleic acid on amino acid acceptor activity and ribosomal binding. J Biol Chem *243*, 1277-1282.

Molinie, B., Wang, J., Lim, K.S., Hillebrand, R., Lu, Z.X., Van Wittenberghe, N., Howard, B.D., Daneshvar, K., Mullen, A.C., Dedon, P.*, et al.* (2016). m(6)A-LAIC-seq reveals the census and complexity of the m(6)A epitranscriptome. Nat Methods *13*, 692-698.

Moreau, C., Ashamu, G.A., Bailey, V.C., Galione, A., Guse, A.H., and Potter, B.V. (2011). Synthesis of cyclic adenosine 5'-diphosphate ribose analogues: a C2'endo/syn "southern" ribose conformation underlies activity at the sea urchin cADPR receptor. Org Biomol Chem *9*, 278-290.

Motorin, Y., Muller, S., Behm-Ansmant, I., and Branlant, C. (2007). Identification of modified residues in RNAs by reverse transcription-based methods. Methods Enzymol *425*, 21-53.

Mueller, E.G., Palenchar, P.M., and Buck, C.J. (2001). The role of the cysteine residues of ThiI in the generation of 4-thiouridine in tRNA. J Biol Chem *276*, 33588-33595.

Murray, J.M., Watson, A.T., and Carr, A.M. (2016). Transformation of Schizosaccharomyces pombe: Lithium Acetate/ Dimethyl Sulfoxide Procedure. Cold Spring Harb Protoc *2016*, pdb prot090969.

Musken, M., Di Fiore, S., Dotsch, A., Fischer, R., and Haussler, S. (2010). Genetic determinants of Pseudomonas aeruginosa biofilm establishment. Microbiology *156*, 431-441.

Noon, K.R., Guymon, R., Crain, P.F., McCloskey, J.A., Thomm, M., Lim, J., and Cavicchioli, R. (2003). Influence of temperature on tRNA modification in archaea: Methanococcoides burtonii (optimum growth temperature [Topt], 23 degrees C) and Stetteria hydrogenophila (Topt, 95 degrees C). J Bacteriol *185*, 5483-5490.

Novoa, E.M., Mason, C.E., and Mattick, J.S. (2017). Charting the unknown epitranscriptome. Nat Rev Mol Cell Biol *18*, 339-340.

O'Connor, M., Lee, W.M., Mankad, A., Squires, C.L., and Dahlberg, A.E. (2001). Mutagenesis of the peptidyltransferase center of 23S rRNA: the invariant U2449 is dispensable. Nucleic Acids Res *29*, 710-715.

Ogle, J.M., Brodersen, D.E., Clemons, W.M., Jr., Tarry, M.J., Carter, A.P., and Ramakrishnan, V. (2001). Recognition of cognate transfer RNA by the 30S ribosomal subunit. Science *292*, 897-902.

Pan, D., Qin, H., and Cooperman, B.S. (2009). Synthesis and functional activity of tRNAs labeled with fluorescent hydrazides in the D-loop. RNA *15*, 346-354.

Panwar, B., and Raghava, G.P. (2014). Prediction of uridine modifications in tRNA sequences. BMC Bioinformatics *15*, 326.

Park, F., Gajiwala, K., Noland, B., Wu, L., He, D., Molinari, J., Loomis, K., Pagarigan, B., Kearins, P., Christopher, J.*, et al.* (2004). The 1.59 A resolution crystal structure of TM0096, a flavin mononucleotide binding protein from Thermotoga maritima. Proteins *55*, 772-774.

Patil, A., Dyavaiah, M., Joseph, F., Rooney, J.P., Chan, C.T., Dedon, P.C., and Begley, T.J. (2012). Increased tRNA modification and gene-specific codon usage regulate cell cycle progression during the DNA damage response. Cell Cycle *11*, 3656-3665.

Patil, D.P., Chen, C.K., Pickering, B.F., Chow, A., Jackson, C., Guttman, M., and Jaffrey, S.R. (2016). m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature *537*, 369-373.

Peng, W.T., Robinson, M.D., Mnaimneh, S., Krogan, N.J., Cagney, G., Morris, Q., Davierwala, A.P., Grigull, J., Yang, X., Zhang, W.*, et al.* (2003). A panoramic view of yeast noncoding RNA processing. Cell *113*, 919-933.

Phizicky, E.M., Grayhack, E.J., Chernyakov, I., and Whipple, J.M. (2009). Roles of tRNA modifications in tRNA turnover. DNA and RNA modification enzymes: structure, mechanism, function and evolution. Landes Bioscience, 564-576.

Phizicky, E.M., and Hopper, A.K. (2010). tRNA biology charges to the front. Genes Dev *24*, 1832-1860.

Pomerantz, S.C., and McCloskey, J.A. (1990). Analysis of RNA hydrolyzates by liquid chromatography-mass spectrometry. Methods Enzymol *193*, 796-824.

Popova, A.M., and Williamson, J.R. (2014). Quantitative analysis of rRNA modifications using stable isotope labeling and mass spectrometry. J Am Chem Soc *136*, 2058-2069.

Quigley, G.J., and Rich, A. (1976). Structural domains of transfer RNA molecules. Science *194*, 796-806.

Quinn, J.J., and Chang, H.Y. (2016). Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet *17*, 47-62.

Ralston, A. (2008). Operons and prokaryotic gene regulation. Nature Education *1*, 216.

Randerath, K., Gupta, R.C., and Randerath, E. (1980). 3H and 32P derivative methods for base composition and sequence analysis of RNA. Methods Enzymol *65*, 638-680.

Reimer, M.L., Schram, K.H., Nakano, K., and Yasaka, T. (1989). The identification of 5,6-dihydrouridine in normal human urine by combined gas chromatography/mass spectrometry. Anal Biochem *181*, 302-308.

Rider, L.W., Ottosen, M.B., Gattis, S.G., and Palfey, B.A. (2009). Mechanism of dihydrouridine synthase 2 from yeast and the importance of modifications for efficient tRNA reduction. J Biol Chem *284*, 10324-10333.

Rietveld, K., Van Poelgeest, R., Pleij, C.W., Van Boom, J.H., and Bosch, L. (1982). The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA. Nucleic Acids Res *10*, 1929-1946.

Rohrer, D.C., and Sundaralingam, M. (1970). Stereochemistry of nucleic acids and their constituents. VI. The crystal structure and conformation of dihydrouracil: a minor base of transfer-ribonucleic acid. Acta Crystallogr B *26*, 546-553.

Rose, R.E., Pazos, M.A., 2nd, Curcio, M.J., and Fabris, D. (2016). Global Epitranscriptomics Profiling of RNA Post-Transcriptional Modifications as an Effective Tool for Investigating the Epitranscriptomics of Stress Response. Mol Cell Proteomics *15*, 932-944.

Rottman, F., and Cerutti, P. (1966). Template activity of uridylic acid-dihydrouridylic acid copolymers. Proc Natl Acad Sci U S A *55*, 960-966.

Roundtree, I.A., Evans, M.E., Pan, T., and He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. Cell *169*, 1187-1200.

Rowland, P., Bjornberg, O., Nielsen, F.S., Jensen, K.F., and Larsen, S. (1998). The crystal structure of Lactococcus lactis dihydroorotate dehydrogenase A complexed with the enzyme reaction product throws light on its enzymatic function. Protein Sci *7*, 1269-1279.

Royburman, P., Royburman, S., and Visser, D.W. (1965). Incorporation of 5,6-dihydrouridine triphosphate into ribonucleic acid by DNA-dependent RNA polymerase. Biochem Biophys Res Commun *20*, 291-297.

Ryvkin, P., Leung, Y.Y., Silverman, I.M., Childress, M., Valladares, O., Dragomir, I., Gregory, B.D., and Wang, L.S. (2013). HAMR: high-throughput annotation of modified ribonucleotides. RNA *19*, 1684-1692.

Saenger, W. (1984). Principles of nucleic acid structure. Springer Advanced Texts in Chemistry.

Safra, M., Sas-Chen, A., Nir, R., Winkler, R., Nachshon, A., Bar-Yaacov, D., Erlacher, M., Rossmanith, W., Stern-Ginossar, N., and Schwartz, S. (2017). The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. Nature *551*, 251-255.

Sakurai, M., Yano, T., Kawabata, H., Ueda, H., and Suzuki, T. (2010). Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. Nat Chem Biol *6*, 733-740.

Saletore, Y., Meyer, K., Korlach, J., Vilfan, I.D., Jaffrey, S., and Mason, C.E. (2012). The birth of the Epitranscriptome: deciphering the function of RNA modifications. Genome Biol *13*, 175.

Savage, D.F., de Crecy-Lagard, V., and Bishop, A.C. (2006). Molecular determinants of dihydrouridine synthase activity. FEBS Lett *580*, 5198-5202.

Schaefer, M., Kapoor, U., and Jantsch, M.F. (2017). Understanding RNA modifications: the promises and technological bottlenecks of the 'epitranscriptome'. Open Biol *7*.

Schaefer, M., Pollex, T., Hanna, K., and Lyko, F. (2009). RNA cytosine methylation analysis by bisulfite sequencing. Nucleic Acids Res *37*, e12.

Schaffrath, R., and Leidel, S.A. (2017). Wobble uridine modifications-a reason to live, a reason to die?! RNA Biol *14*, 1209-1222.

Schleich, H.G., Wintermeyer, W., and Zachau, H.G. (1978). Replacement of wybutine by hydrazines and its effect on the active conformation of yeast tRNAPhe. Nucleic Acids Res *5*, 1701-1713.

Schnackerz, K.D., Dobritzsch, D., Lindqvist, Y., and Cook, P.F. (2004). Dihydropyrimidine dehydrogenase: a flavoprotein with four iron-sulfur clusters. Biochim Biophys Acta *1701*, 61-74.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat Methods *9*, 671-675.

Schwartz, S., Agarwala, S.D., Mumbach, M.R., Jovanovic, M., Mertins, P., Shishkin, A., Tabach, Y., Mikkelsen, T.S., Satija, R., Ruvkun, G.*, et al.* (2013). High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell *155*, 1409-1421.

Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., Leon-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S.*, et al.* (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell *159*, 148-162.

Schwartz, S., and Motorin, Y. (2017). Next-generation sequencing technologies for detection of modified nucleotides in RNAs. RNA Biol *14*, 1124-1137.

Seelam, P.P., Sharma, P., and Mitra, A. (2017). Structural landscape of base pairs containing post-transcriptional modifications in RNA. RNA *23*, 847-859.

Sharma, S., and Lafontaine, D.L.J. (2015). 'View From A Bridge': A New Perspective on Eukaryotic rRNA Base Modification. Trends Biochem Sci *40*, 560-575.

Shigi, N., Suzuki, T., Terada, T., Shirouzu, M., Yokoyama, S., and Watanabe, K. (2006). Temperature-dependent biosynthesis of 2-thioribothymidine of Thermus thermophilus tRNA. J Biol Chem *281*, 2104-2113.

Shih, T.Y., and Bonner, J. (1969). Chromosomal RNA of calf thymus chromatin. Biochim Biophys Acta *182*, 30-35.

Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M*., et al.* (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. Nat Methods *12*, 323-325.

Sipa, K., Sochacka, E., Kazmierczak-Baranska, J., Maszewska, M., Janicka, M., Nowak, G., and Nawrot, B. (2007). Effect of base modifications on structure, thermodynamic stability, and gene silencing activity of short interfering RNA. RNA *13*, 1301-1316.

Sloan, K.E., Warda, A.S., Sharma, S., Entian, K.D., Lafontaine, D.L.J., and Bohnsack, M.T. (2017). Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. RNA Biol *14*, 1138-1152.

Smrt, J., Kemper, W., Caskey, T., and Nirenberg, M. (1970). Template activity of modified terminator codons. J Biol Chem *245*, 2753-2757.

Smrt, J., Skoda, J., Lisy, V., and Sorm, F. (1966). Loss of coding properties of the trinucleotide guanylyl-uridylyl-uridine on replacement of uridylic by dihydrouridylic acid. Biochim Biophys Acta *129*, 210-211.

Song, J., and Yi, C. (2017). Chemical Modifications to RNA: A New Layer of Gene Expression Regulation. ACS Chem Biol *12*, 316-325.

Squires, J.E., Patel, H.R., Nousch, M., Sibbritt, T., Humphreys, D.T., Parker, B.J., Suter, C.M., and Preiss, T. (2012). Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res *40*, 5023-5033.

Sridharan, G., Ramani, P., and Patankar, S. (2017). Serum metabolomics in oral leukoplakia and oral squamous cell carcinoma. J Cancer Res Ther *13*, 556-561.

Stuart, J.W., Basti, M.M., Smith, W.S., Forrest, B., Guenther, R., Sierzputowska-Gracz, H., Nawrot, B., Malkiewicz, A., and Agris, P.F. (1996). Structure of the trinucleotide D-acp3U-A with coordinated Mg2+ demonstrates that modified nucleosides contribute to regional conformations of RNA. Nucleosides & Nucleotides *15*, 1009-1028.

Suck, D., Saenger, W., and Zechmeister, K. (1971). Conformation of the tRNA minor constituent dihydrouridine. FEBS Lett *12*, 257-259.

Suck, D., Saenger, W., and Zechmeister, K. (1972). Molecular and crystal structure of the tRNA minor constituent dihydrouridine. Acta Cryst *B28*, 596-605.

Sun, Q., Hao, Q., and Prasanth, K.V. (2017). Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression. Trends Genet *34*, 142-157.

Sundaralingam, M., Rao, S.T., and Abola, J. (1971a). Molecular conformation of dihydrouridine: puckered base nucleoside of transfer RNA. Science *172*, 725-727.

Sundaralingam, M., Rao, S.T., and Abola, J. (1971b). Stereochemistry of nucleic acids and their constituents. 23. Crystal and molecular structure of dihydrouridine "hemihydrate," a rare nucleoside with a saturated base occurring in the dihydrouridine loop of transfer ribonucleic acids. J Am Chem Soc *93*, 7055-7062.

Suzuki, T., and Suzuki, T. (2014). A complete landscape of post-transcriptional modifications in mammalian mitochondrial tRNAs. Nucleic Acids Res *42*, 7346-7357.

Tisseur, M., Kwapisz, M., and Morillon, A. (2011). Pervasive transcription - Lessons from yeast. Biochimie *93*, 1889-1896.

Topp, H., Duden, R., and Schoch, G. (1993). 5,6-Dihydrouridine: a marker ribonucleoside for determining whole body degradation rates of transfer RNA in man and rats. Clin Chim Acta *218*, 73-82.

Urbonavicius, J., Qian, Q., Durand, J.M., Hagervall, T.G., and Bjork, G.R. (2001). Improvement of reading frame maintenance is a common function for several tRNA modifications. EMBO J *20*, 4863-4873.

Vandivier, L.E., Campos, R., Kuksa, P.P., Silverman, I.M., Wang, L.S., and Gregory, B.D. (2015). Chemical Modifications Mark Alternatively Spliced and Uncapped Messenger RNAs in Arabidopsis. Plant Cell *27*, 3024-3037.

Vare, V.Y., Eruysal, E.R., Narendran, A., Sarachan, K.L., and Agris, P.F. (2017). Chemical and Conformational Diversity of Modified Nucleosides Affects tRNA Structure and Function. Biomolecules *7*.

Vogeli, G. (1979). The nucleotide sequence of tRNA tyrosine from the fission yeast Schizosaccharomyces pombe. Nucleic Acids Res *7*, 1059-1065.

Wang, T., Shigdar, S., Shamaileh, H.A., Gantier, M.P., Yin, W., Xiang, D., Wang, L., Zhou, S.F., Hou, Y., Wang, P.*, et al.* (2017). Challenges and opportunities for siRNA-based cancer treatment. Cancer Lett *387*, 77-83.

Westhof, E., Dumas, P., and Moras, D. (1985). Crystallographic refinement of yeast aspartic acid transfer RNA. J Mol Biol *184*, 119-145.

Westhof, E., and Sundaralingam, M. (1986). Restrained refinement of the monoclinic form of yeast phenylalanine transfer RNA. Temperature factors and dynamics, coordinated waters, and base-pair propeller twist angles. Biochemistry *25*, 4868-4878.

Whelan, F., Jenkins, H.T., Griffiths, S.C., Byrne, R.T., Dodson, E.J., and Antson, A.A. (2015). From bacterial to human dihydrouridine synthase: automated structure determination. Acta Crystallogr D Biol Crystallogr *71*, 1564-1571.

Wilusz, J.E., Freier, S.M., and Spector, D.L. (2008). 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell *135*, 919-932.

Wilusz, J.E., Whipple, J.M., Phizicky, E.M., and Sharp, P.A. (2011). tRNAs marked with CCACCA are targeted for degradation. Science *334*, 817-821.

Wintermeyer, W., Schleich, H.G., and Zachau, H.G. (1979). Incorporation of amines or hydrazines into tRNA replacing wybutine or dihydrouracil. Methods Enzymol *59*, 110-121.

Wintermeyer, W., and Zachau, H.G. (1970). A specific chemical chain scission of tRNA at 7-methylguanosine. FEBS Lett *11*, 160-164.

Wintermeyer, W., and Zachau, H.G. (1971). Replacement of Y base, dihydrouracil, and 7-methylguanine in tRNA by artificial odd bases. FEBS Lett *18*, 214-218.

Wintermeyer, W., and Zachau, H.G. (1974). Replacement of odd bases in tRNA by fluorescent dyes. Methods Enzymol *29*, 667-673.

Wintermeyer, W., and Zachau, H.G. (1979). Fluorescent derivatives of yeast tRNAPhe. Eur J Biochem *98*, 465-475.

Wong, T.W., McCutchan, T., Kohli, J., and Soll, D. (1979). The nucleotide sequence of the major glutamate transfer RNA from Schizosaccharomyces pombe. Nucleic Acids Res *6*, 2057-2068.

Woo, N.H., Roe, B.A., and Rich, A. (1980). Three-dimensional structure of Escherichia coli initiator tRNAfMet. Nature *286*, 346-351.

Xing, F., Hiley, S.L., Hughes, T.R., and Phizicky, E.M. (2004). The specificities of four yeast dihydrouridine synthases for cytoplasmic tRNAs. J Biol Chem *279*, 17850-17860.

Xing, F., Martzen, M.R., and Phizicky, E.M. (2002). A conserved family of Saccharomyces cerevisiae synthases effects dihydrouridine modification of tRNA. RNA *8*, 370-381.

Yang, X., Yang, Y., Sun, B.F., Chen, Y.S., Xu, J.W., Lai, W.Y., Li, A., Wang, X., Bhattarai, D.P., Xiao, W.*, et al.* (2017). 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. Cell Res *27*, 606-625.

Yokoyama, S., Watanabe, K., and Miyazawa, T. (1987). Dynamic structures and functions of transfer ribonucleic acids from extreme thermophiles. Adv Biophys *23*, 115-147.

You, C., Dai, X., and Wang, Y. (2017). Position-dependent effects of regioisomeric methylated adenine and guanine ribonucleosides on translation. Nucleic Acids Res *45*, 9059-9067.

Yu, F., Tanaka, Y., Yamashita, K., Suzuki, T., Nakamura, A., Hirano, N., Suzuki, T., Yao, M., and Tanaka, I. (2011). Molecular basis of dihydrouridine formation on tRNA. Proc Natl Acad Sci U S A *108*, 19593-19598.

Zallot, R., Brochier-Armanet, C., Gaston, K.W., Forouhar, F., Limbach, P.A., Hunt, J.F., and de Crecy-Lagard, V. (2014). Plant, animal, and fungal micronutrient queuosine is salvaged by members of the DUF2419 protein family. ACS Chem Biol *9*, 1812-1825.

Zhang, Y., Buchholz, F., Muyrers, J.P., and Stewart, A.F. (1998). A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet *20*, 123-128.

Zhao, K., Liu, M., and Burgess, R.R. (2005). The global transcriptional response of Escherichia coli to induced sigma 32 protein involves sigma 32 regulon activation followed by inactivation and degradation of sigma 32 in vivo. J Biol Chem *280*, 17758-17768.

ANNEXES

I. ABBREVIATIONS

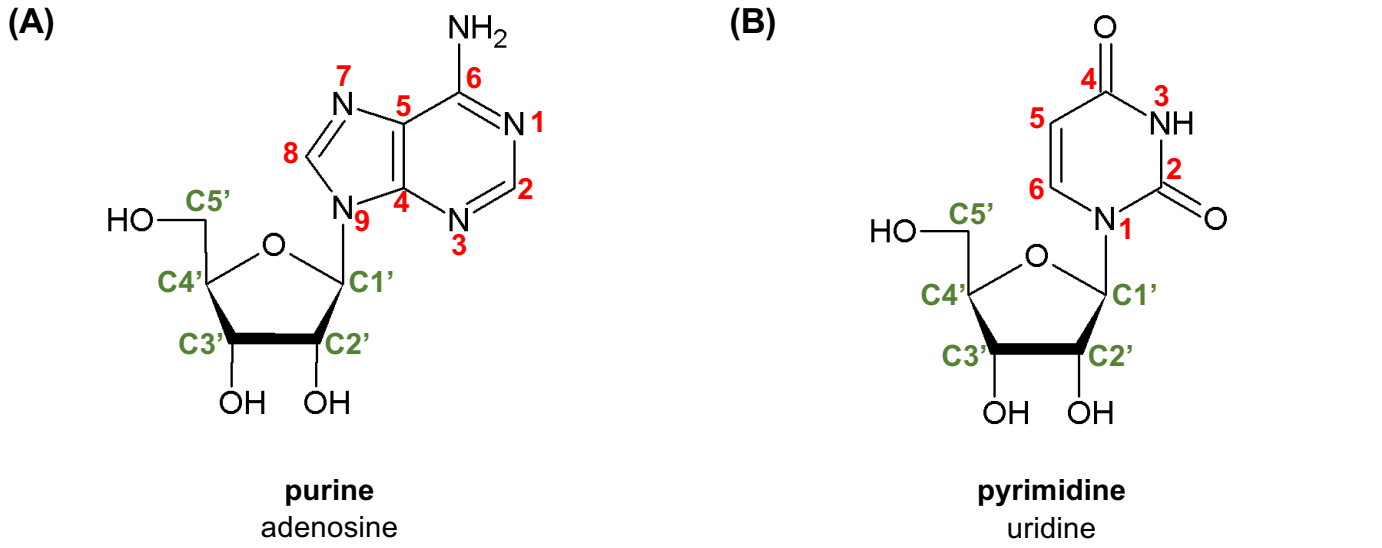
ψ	pseudouridine
Δ*3*	*E. coli* triple *dusA-B-C* mutant
Δ*4*	fission yeast quadruple *dus1-2-3-4* mutant
5'-3'	nucleic acids polarity
A	adenosine
aa	amino acid
Ab	antibody
ac^4C	N^4-acetylcytidine
acp^3D	3-(3-amino-3-carboxypropyl)-5,6-dihydrouridine
acp^3U	3-(3-amino-3-carboxypropyl)uridine
Ala	alanine
Arg	arginine
ASL	tRNA anticodon stem and loop domain
Asn	asparagine
Asp	aspartic acid
bp	base pair
C	carbon
cDNA	complementary DNA
CDS	coding DNA sequence
CMC	N-cyclohexyl-N′-β-(4-methylmorpholinium)ethylcarbodiimide
$cmnm^5s^2U$	5-carboxymethylaminomethyl-2-thiouridine
crRNA	CRISPR RNA
Cys(-SH)	cysteine
D	dihydrouridine
D-seq	dihydrouridine sequencing
ddC	dideoxycytidine
dihydro-UMP	dihydrouridine monophosphate or dihydrouridylate
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DSL	tRNA D stem and loop domain
dsRBD	double-stranded RNA binding motif
DSRM	double-stranded RNA binding motif
dus or *DUS*	dihydrouridine synthase gene
Dus or DUS	dihydrouridine synthase protein
EtBr	ethidium bromide
fc	fold change
FDR	false discovery rate
FMN	flavin mononucleotide
G	guanosine
Gly	glycine
Gm	2'-O-methylguanosine
GST	glutathione-S-transferase
HIV	human immunodeficiency virus

hm^5C	5-hydroxymethylcytidine
*Hs*Δ*4*	HCT 116 cell line quadruple *DUS1-2-3-4* mutant
I	inosine
IP	immunoprecipitation
kb	kilobase
KO	knockdown
LC/MS	high performance liquid chromatography-mass spectrometry
Leu	leucine
lincRNA	long intergenic noncoding RNA
lncRNA	long noncoding RNA
Lys	lysine
m^1A	1-methyladenosine
$m^1acp^3\psi$	1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine
m^1G	1-methylguanosine
$m^{2,2}G$	N^2,N^2-dimethylguanosine
m^2G	N^2-methylguanosine
m^5C	5-methylcytidine
m^5s^2U	5-methyl-2-thiouridine or 2-thioribothymidine
m^5U	5-methyluridine
$m^{6,6}A$	N^6,N^6-dimethyladenosine
m^6A	N^6-methyladenosine
m^6Am	N^6-2'-O-dimethyladenosine
m^7G	7-methylguanosine
MALAT1	metastasis-associated lung adenocarcinoma transcript 1
Mbp	megabase pairs
mcm^5s^2U	5-methoxycarbonylmethyl-2-thiouridine
miRNA	microRNA
mnm^5s^2U	5-methylaminomethyl-2-thiouridine
mRNA	messenger RNA
mt-tRNA	mitochondrial tRNA
N	nitrogen
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
ncRNA	noncoding RNA
NGS	next-generation sequencing
nm	nanometer
NMR	nuclear magnetic resonance
nt	nucleotide
OD	optical density
ORF	open reading frame
p	5'-3' phosphodiester bond
Phe	phenylalanine
pmol	picomole
pre-mRNA	precursor-messenger RNA
pre-rRNA	precursor-ribosomal RNA
pre-tRNA	precursor-transfer RNA
pre-$tRNA_i^{Met}$	precursor of initiator methionine transfer RNA
PRF	programmed ribosomal frameshifting
PTM	RNA posttranscriptional modification
Q	queuosine

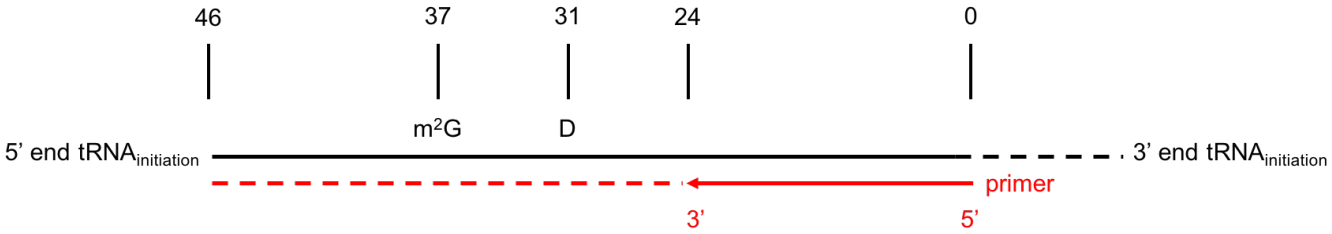
QC	quality control
qRT-PCR	quantitative reverse transcription-polymerase chain reaction
RBP	RNA binding protein
RNA	ribonucleic acid
RNA-seq	RNA sequencing
rRNA	ribosomal RNA
RT	reverse transcription or room temperature
RTase	reverse transcriptase
RTD	rapid tRNA decay
s^4U	4-thiouridine
SEM	standard error of mean
Ser	serine
snoRNA	small nucleolar RNA
snoRNP	small nucleolar ribonucleoprotein
snRNA	small nuclear RNA
sp.	species
T7P	promoter sequence for the T7 RNA polymerase
tDNA	DNA sequence of a tRNA
THC	tetrahydrocytidine
THU	tetrahydrouridine
tiRNA	tRNA-derived stress induced RNA
TLC	thin layer chromatography
TLD	tRNA-like domain
tmRNA	transfer-messenger RNA
tracrRNA	transactivating crRNA
tRF	tRNA-derived fragment
tRNA	transfer RNA
ts	temperature-sensitive
TSL	tRNA T stem and loop domain
tsRNA	tRNA-derived small RNA
Tyr	tyrosine
U	uridine
UTR	untranslated region
UV	ultraviolet
Val	valine
vs	versus
yW	wybutosine

II. SUPPLEMENTAL INFORMATION

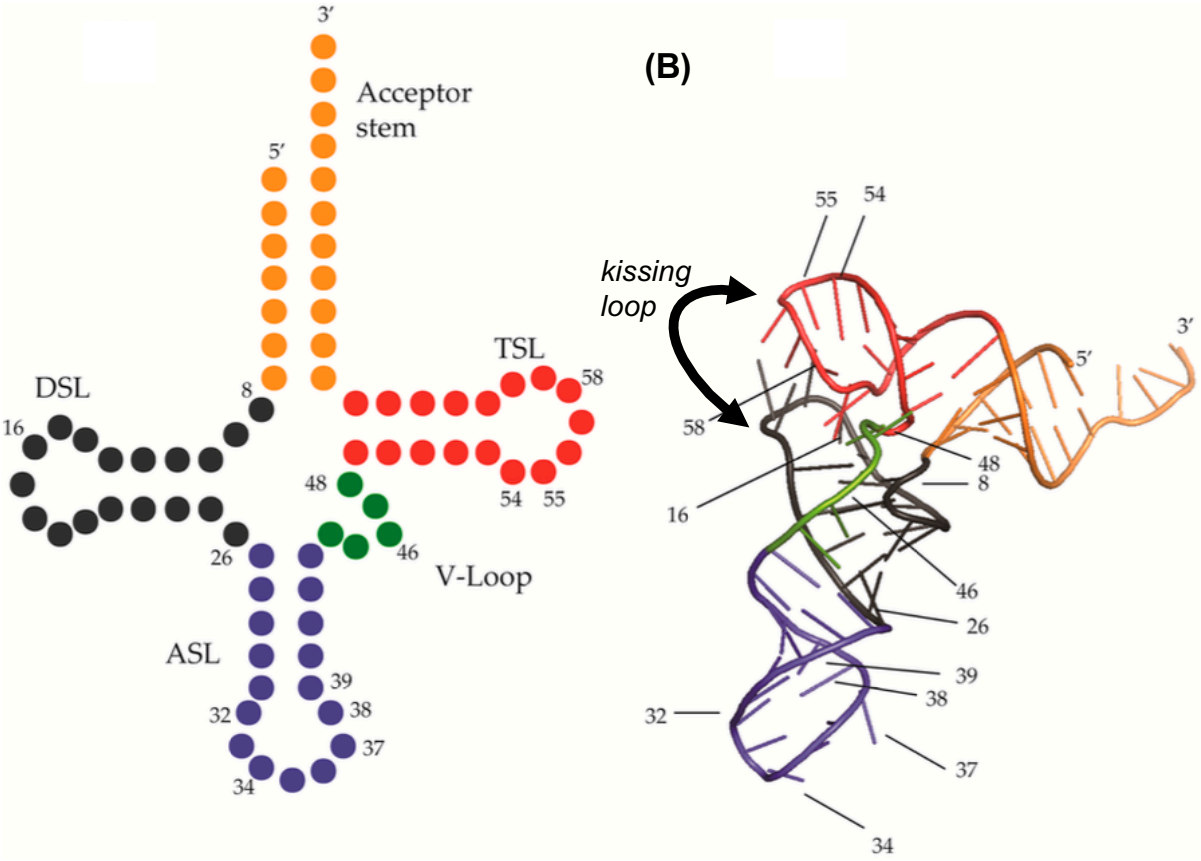
The supplemental figures (S1 to S19) are found on the following pages. The Tables S1, S2 and S3 are at the end of this section.



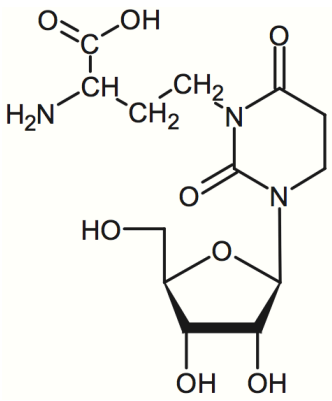
Supplemental Figure 1 – **Nucleosides numbering system**. **(A)** The adenosine ribonucleoside (a purine) has nine numbered atoms on its nucleobase (red) and five numbered carbons on its ribose (green). (B) The uridine ribonucleoside (a pyrimidine) has six numbered atoms on its nucleobase (red) and the annotation for the ribose remains unchanged (Boccaletto et al., 2018).



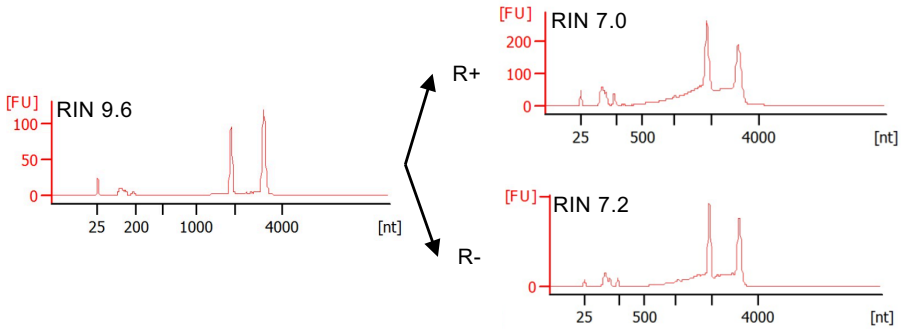
Supplemental Figure 2 – **Principle of primer extension, $tRNA_{initiation}$ of *Schizosaccharomyces pombe***. A radiolabeled primer (solid red line) is added into a denatured sample of total RNA. The sequence of the primer is complementary to the RNA molecule of interest. After hybridization, the reverse transcription is performed and the product of this reaction (solid and dashed red lines) is loaded on a polyacrylamide gel. The detection of the radiolabed polynucleotides of different sizes corresponds to the positions where the reverse transcriptase was stopped or blocked. In this specific example, the reaction is stopped 22 nucleotides upstream of the 3'-end of the 24nt-long radiolabeled primer (22 + 24 = 46nt-long polynucleotide). The N^2-methylguanosine (m^2G) disturbs the Watson-Crick interaction and is therefore a naturally-blocking modification for the reverse transcriptase. Finally, the dihydrouridine (D) is positioned 31 nucleotides upstream of the 5'-end of the primer. After R+ treatment (see Results and Discussion), the rhodamine incorporated at the D position blocks the reverse transcription reaction one nucleotide downstream of the D position, resulting in a 30nt-long radiolabeled polynucleotide.



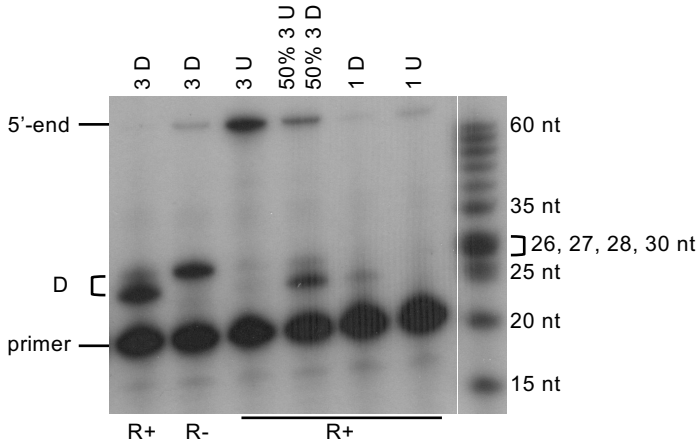
Supplemental Figure 3 – **Secondary (cloverleaf) (A) and tertiary (L-shaped) (B) structures of a tRNA**. DSL in black (D stem and loop domain), ASL in purple (anticodon stem and loop domain), V-loop in green (variable loop domain), TSL in red (T stem and loop domain), acceptor stem in orange. Numbers indicate tRNA positions (anticodon = 34-35-36) (Vare et al., 2017).



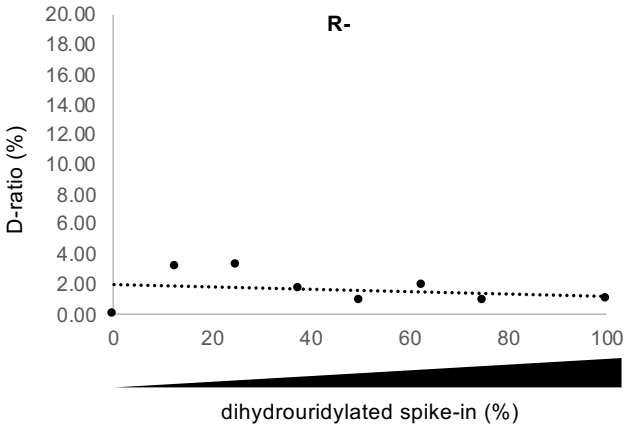
Supplemental Figure 4 – **Chemical structure of acp^{3}D (Krog et al., 2011).**



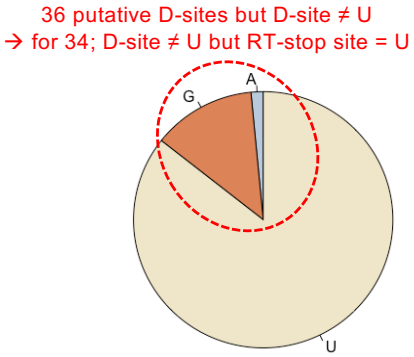
Supplemental Figure 5 – **Microchip analysis of yeast total RNA.** The Rho labeling on total RNA extracted from WT yeast moderately degrades the RNA as observed by a decrease of the RIN values (0 > RNA Integrity Number > 10). FU (fluorescence units), nt (length expressed in nucleotides).



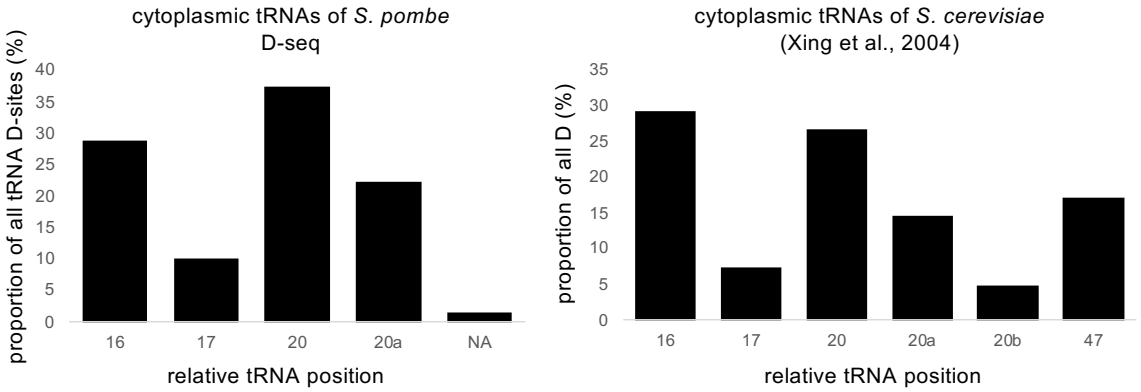
Supplemental Figure 6 – **Effect of three consecutive dihydrouridines on RT termination.** Primer extension assay performed on an *in vitro* transcribed RNA mixed with *S. pombe* total RNA before R+ or R- chemical reactions. 3 D indicates an *in vitro* RNA containing three consecutive D residues that could virtually generate three truncated polynucleotides after the RT in the R+ condition at 25, 26 and 27nt, respectively. 3 U indicates an *in vitro* transcribed RNA without dihydrouridylated residues (there are three uridines instead). 1 D and 1 U are synthetic RNAs as in Fig. 10F with a unique (dihydro)uridine.



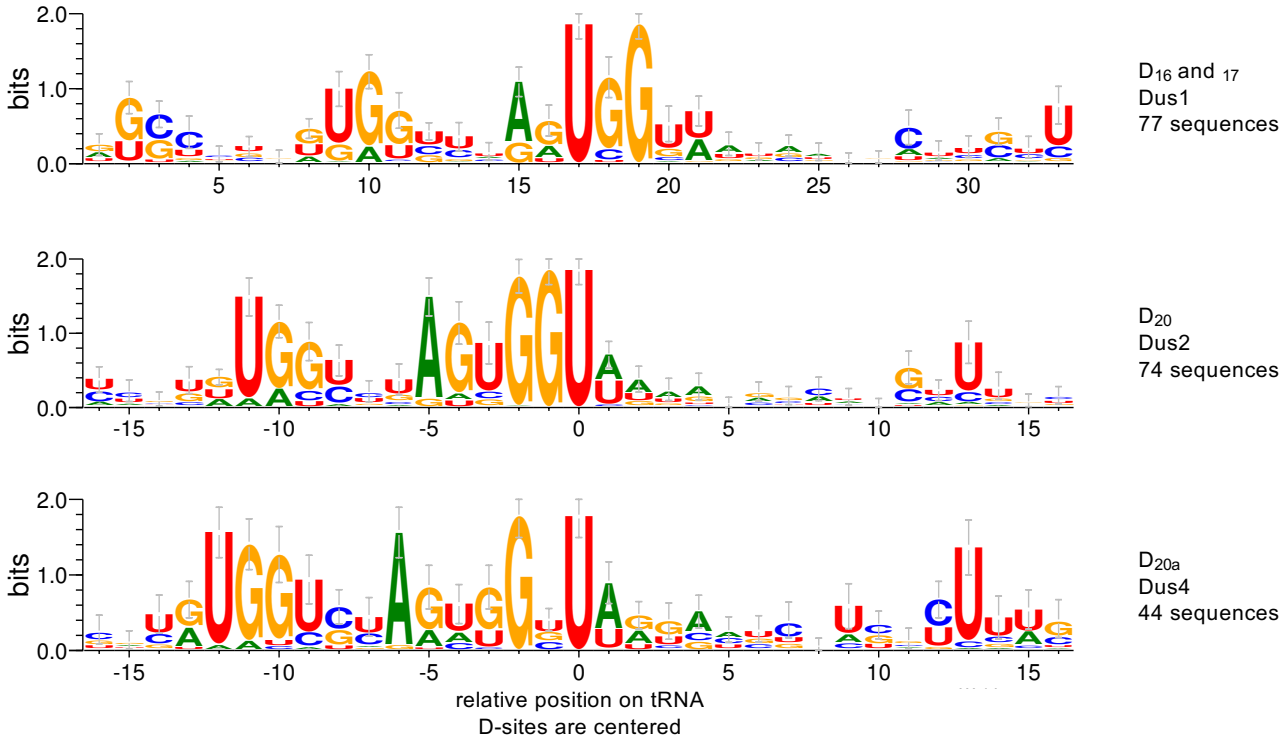
Supplemental Figure 7 – D-ratio of the spike-in position 45 does not gradually increase along with D:U content in the R- condition. 100% means that all spike-in RNAs added to the sample carry a single D whereas 0% corresponds to a U spike-in.



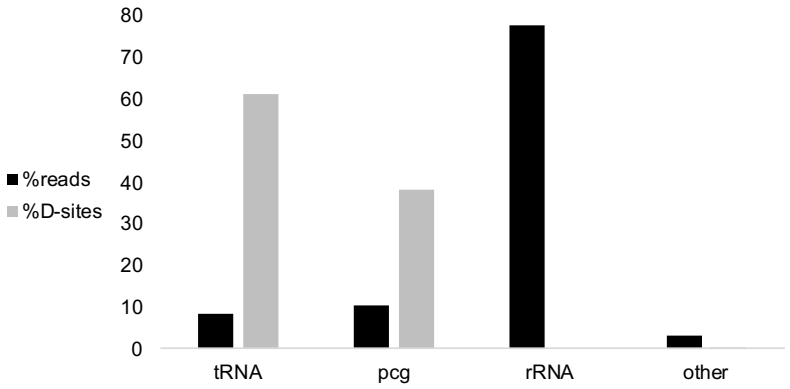
Supplemental Figure 8 – Effect of the '*RT-stop site upstream nucleotide = U*' condition on the detection of D on yeast tRNAs. Getting rid of the last condition of the D-seq pre-filtering analysis (Fig. 12) led to the detection of 36 new D-sites on yeast tRNAs (for a total of 264 putative D). 94.4% of these new sites were not downstream of a U but were U themselves.



Supplemental Figure 9 – Comparison of the dihydrouridylated positions on tRNAs between fission and budding yeasts. On the left, the 198 cytoplasmic tRNA D-sites are manually curated and assigned to a canonical D positions (D_{16}, $_{17}$, $_{20}$, $_{20a}$ or $_{47}$) if possible. NA stands for *not assigned* (D_2-tRNA^{ArgCCU}, D_8-tRNA^{MetCAU} and D_{29}-tRNA^{ProCGG}). **On the right,** distribution of the 82 unique cytoplasmic tRNA dihydrouridine positions in the budding yeast as referred by the work of Phizicky and colleagues.



Supplemental Figure 10 – **Seeking for a dihydrouridylation motif in yeast tRNAs.** Consensus sequence surrounding 195 cytoplasmic tRNA D-sites distributed according to the conserved substrate specificities of Dus enzymes.



Supplemental Figure 11 – **Absence of a strict correlation between number of detected D-sites and coverage.** D-seq in fission yeast resulted in 372 detected D-sites that are all represented by the light gray bars. Black bars represent the distribution of sequenced reads according to the RNA species. pcg (protein-coding gene product or mRNA).

Category_Name	GeneSet_Name	Representation	Corrected p-value
GO Molecular Function	oxidoreductase activity, acting on NAD(P)H	Enriched	0.00540684
GO Molecular Function	oxidoreductase activity, acting on a sulfur group of donors, NAD(P) as acceptor	Enriched	0.00545604
Gene Expression	Highly expressed genes	Enriched	0.000199905
Phenotypes (FYPO)	inviable after spore germination, without cell division, with normal germ tube morphology	Enriched	0.000142192
Phenotypes (FYPO)	inviable after spore germination with normal, unseptated germ tube morphology	Enriched	0.000142192
Phenotypes (FYPO)	inviable spore	Enriched	0.00119614
Phenotypes (FYPO)	inviable after spore germination, without cell division	Enriched	0.00809965
Protein Features	Alanine	Higher	2.46E-11
Protein Features	Number of amino acids	Higher	4.65E-08
Protein Features	Protein copies per proliferating cell	Higher	4.65E-08
Protein Features	Molecular weight (kDa)	Higher	2.44E-07
Protein Features	Glycine	Higher	1.96E-06
Protein Features	Charge	Lower	0.000125472
Protein Features	Isoelectric point (predicted pH)	Lower	0.000230274
Protein Features	Nitrogen content	Lower	0.00228755
Transcript Features	mRNA copies per proliferating cell	Higher	1.31E-32
Transcript Features	mRNA level (WT)	Higher	2.45E-22
Transcript Features	mRNA copies per quiescent cell	Higher	1.45E-20
Transcript Features	Relative Pol II occupancy	Higher	2.72E-11
Transcript Features	Annotated transcript length	Higher	3.43E-06
Transcript Features	Ribosomal density (rb/kb)	Lower	0.00086295
Transcript Features	mRNA stabilities	Higher	0.00545604

Supplemental Figure 12 – **Putatively dihydouridylated protein-coding genes enrichment analysis.** A *S. pombe*-specific tool (AnGeLi) was used to determine potential over- or under-representation of the 125 mRNAs D-seq candidates in various categories. Results with a *p*-value smaller than 0.01 are shown.

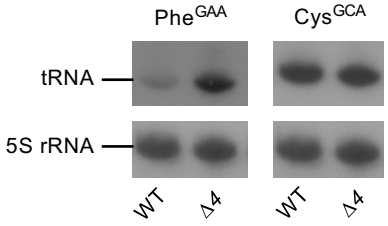
codons	aa
CUA	Leu
CUC	Leu
GUC	Val
UAA	STOP
UAG	STOP
UCA	Ser
UCC	Ser
UCG	Ser
UGA	STOP
UGC	Cys

codons not represented in D-seq (*S. pombe*)
100% have a single U

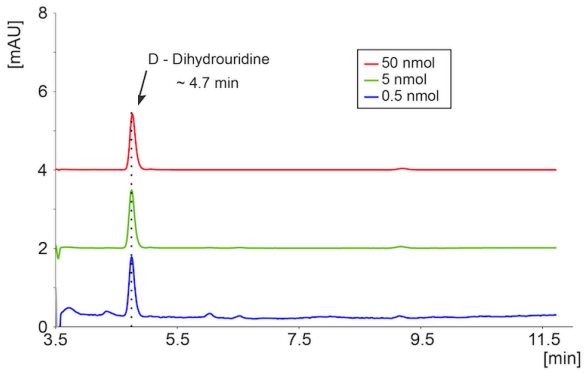
possibilites to have a D-site on these codons = 207,964
→ 8% of the possibilities on the coding transcriptome
→ 0 D-sites

*** $p = 7.3 \times 10^{-5}$

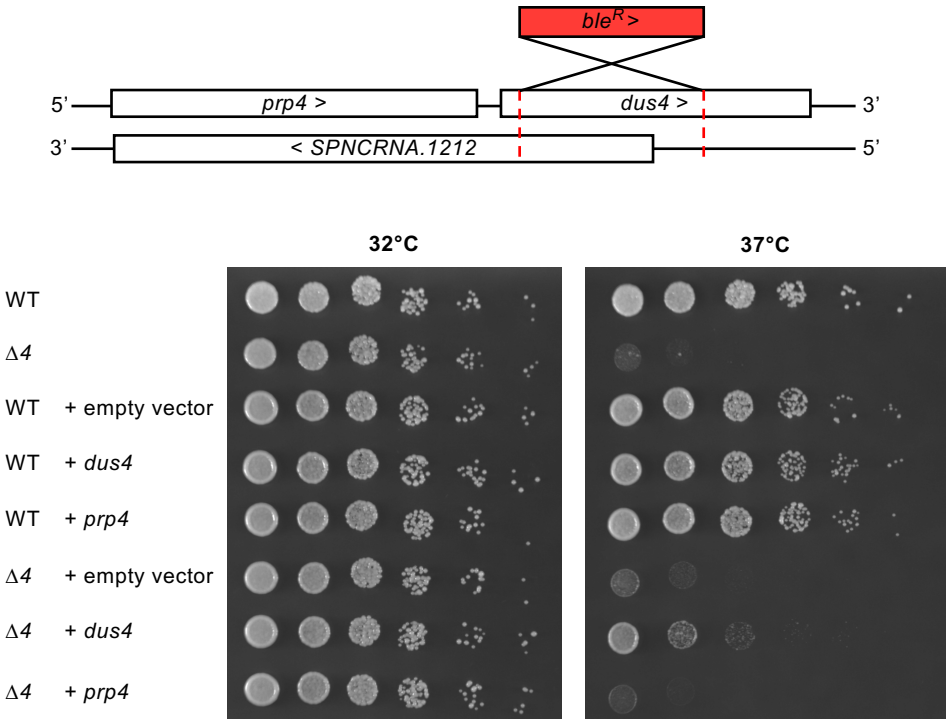
Supplemental Figure 13 – **Non-represented codons in yeast D-seq.** None of the 124 detected D-sites on coding sequences were found on the listed codons although they represent almost one tenth of the U residues found the coding transcriptome (↔ one tenth of the possibilities to get a D-site on the coding transcriptome). One-sided Fisher's Exact test; *** $p < 0.001$.



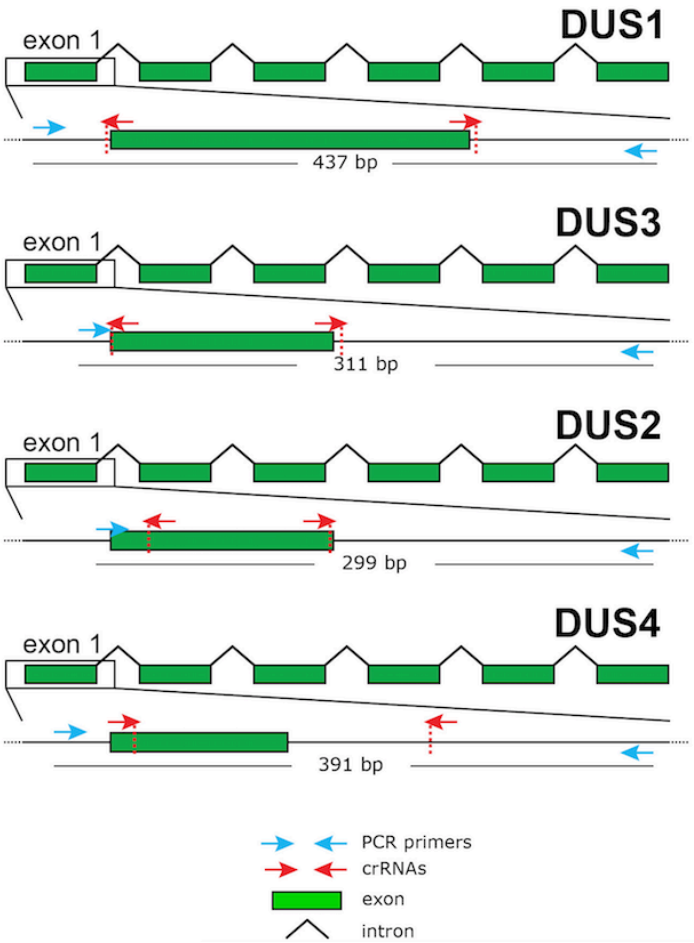
Supplemental Figure 14 – **Hypodihydrouridylated tRNA abundance.** tRNA Northern blotting was performed on two RNA species in WT and *Δ4* yeast strains. 5S rRNA is used as a loading control, representative examples of biological triplicates.



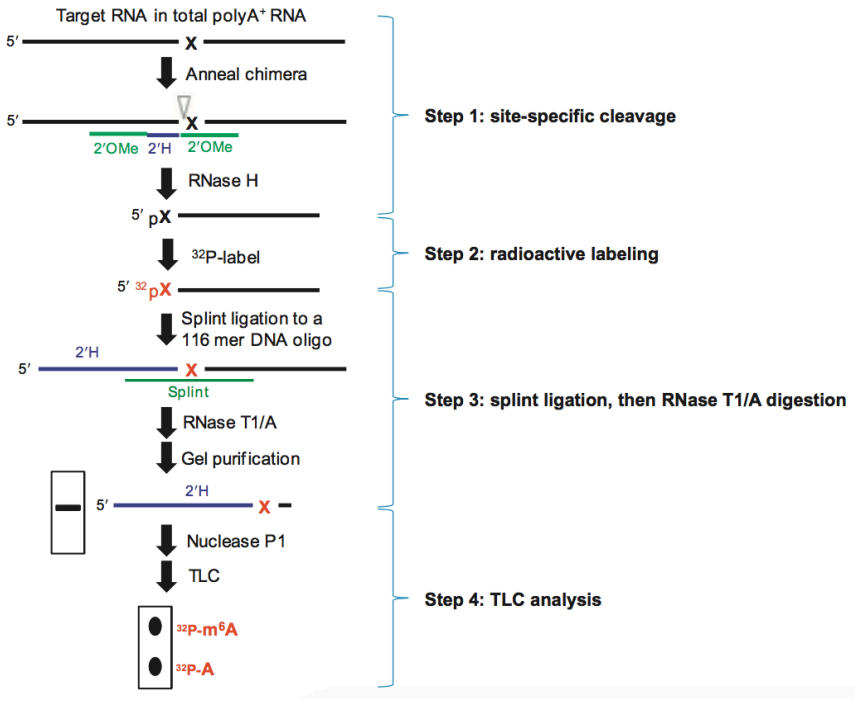
Supplemental Figure 15 – **Calibration of HPLC system for detection of D.** Commercial D nucleoside is efficiently detected and is eluted from the HPLC column after 4.7min. Detection at 247nm. mAU (milli-absorbance units).



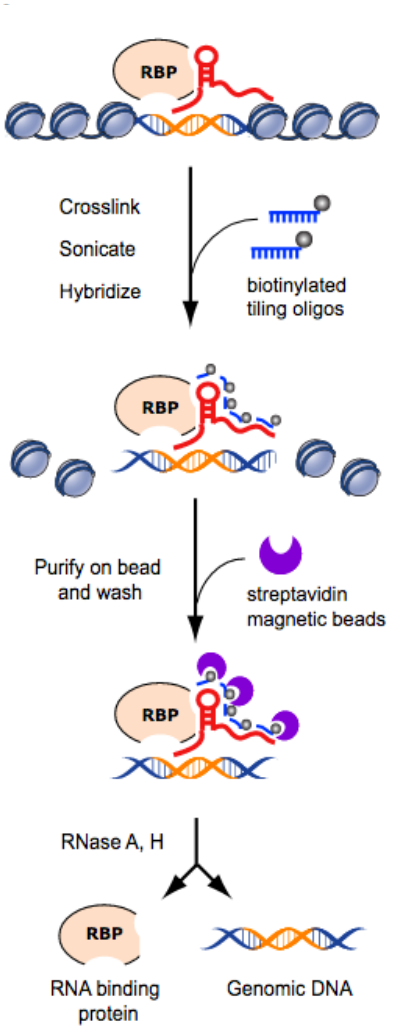
Supplemental Figure 16 – **Yeast quadruple *dus* mutant thermosensitivity is *prp4*-independent. Above:** *dus4* locus; *prp4* is an upstream gene, *SPNCRNA.1212* is an antisense long noncoding RNA. In the *Δ4* genetic background, the *dus4* CDS (delimited by the red dotted lines) was replaced by a phleomycin-resistance cassette gene (*ble*^{R}). As shown in Fig. 19A, *prp4* and *SPNCRNA.1212* expressions were deeply disturbed by this genomic manipulation. **Below:** growth phenotype of the *Δ4* mutant compared to the WT strain. Exponential growing cells were spotted on rich medium and incubated at 32 or 37°C (5-fold dilutions) for 47h. All strains carried an auxotrophic marker for the positive selection of the vector during the (pre-)culturing growth in a defined medium. The vector used in this study resulted in the overexpression of the cloned gene (*dus4* or *prp4*). *prp4* overexpression did not suppress the *Δ4* thermosensitivity whereas *dus4* overexpression partially did so.



Supplemental Figure 17 – **CRISPR-Cas9 strategy implemented for inactivation of human *DUS* genes in the HCT 116 cell line.** Coding exon 1 is targeted for Cas9 mediated-cleavage by two guide RNAs (crRNAs or CRISPR RNAs, red arrows). For each deletion, two ribonucleoproteins (Cas9 + crRNA:transactivating-crRNA duplex) were electroporated, each of them carrying a different crRNA. Deletion diagnostic relied on PCR amplification (primers are depicted by blue arrows and the full-length DNA product size – i.e. in WT context – is indicated for each gene). bp (base pairs).



Suplemental Figure 18 – SCARLET approach for the single-site detection of an RNA modification. RNA of interest (black line) is targeted for a specific cleavage upstream of the putative modified candidate (X). The cleavage is performed by the RNAse H nuclease by means of an RNA-DNA chimeric oligonucleotide (the green line highlights RNA with methylated C2' ribose [2'OMe] and the blue line highlights DNA with deoxygenated ribose [2'H]). 5'-ends are then radioactively labeled with phosphate (^{32}p). Splint ligation with an oligonucleotide (splint, green line) that overlaps the RNA of interest and a synthetic ssDNA (2'H, blue line) is performed, followed by RNA degradation and gel purification. The resulting product is analyzed by one-dimension TLC (Liu and Pan, 2015).



Suplemental Figure 19 – Chromatin isolation by RNA precipitation. ChIRP relies on the *in vivo* crosslinking of an RNA with its interacting partners (DNA or proteins). The RNA of interest is isolated by specific biotinylated oligonucleotides and streptavidin-dependent precipiation. For the determination of RNA-binding proteins, RNA and DNA are degraded before mass spectrometry analysis (Chu et al., 2011).

gene	description	ATG	codon	position
aat1	amino acid transmembrane transporter Aat1	250	UUG	1
abc2	glutathione S-conjugate-exporting ATPase Abc2	4069	UUC	1
ace2	transcription factor Ace2	1392	AUU	3
adg2	conserved fungal protein Adg2	1257	UCU	3
ags1	alpha glucan synthase Ags1	6736	UGG	1
ala1	mitochondrial and cytoplasmic alanine-tRNA ligase Ala1 (predicted)	2430	CAU	3
ala1	mitochondrial and cytoplasmic alanine-tRNA ligase Ala1 (predicted)	1972	UGU	1
ala1	mitochondrial and cytoplasmic alanine-tRNA ligase Ala1 (predicted)	1051	UUC	1
apl4	AP-1 adaptor complex gamma subunit Apl4	2109	GCU	3
arg12	argininosuccinate synthase Arg12	451	UUU	1
bfr1	brefeldin A transmembrane transporter Bfr1	127	UCU	1
bms1	GTP binding protein Bms1 (predicted)	1757	CUU	3
brr2	U5 snRNP complex subunit Brr2	6478	UUU	1
but2	But2 family protein But2	-283	NA	NA
cbr1	cytochrome b5 reductase Cbr1 (predicted)	197	CUU	2
ccr1	NADPH-cytochrome p450 reductase	1070	GUU	2
ccr1	NADPH-cytochrome p450 reductase	167	UUC	2
cct1	chaperonin-containing T-complex alpha subunit Cct1	1829	CUU	2
cct6	chaperonin-containing T-complex zeta subunit Cct6	639	UUU	1
cct7	chaperonin-containing T-complex eta subunit Cct7	1259	UGU	3
cct7	chaperonin-containing T-complex eta subunit Cct7	267	AUU	3
cek1	serine/threonine protein kinase Cek1	3473	GUU	2
cfr1	Chs five related protein Cfr1	846	AAU	3
cip2	RNA-binding protein Cip2	1405	UUU	1
cop1	coatomer alpha subunit Cop1 (predicted)	2298	UAU	3
deb1	transcription factor Deb1/Rdp1	692	CUG	2
def1	RNAPII degradation factor Def1 (predicted)	338	UUG	2
def1	RNAPII degradation factor Def1 (predicted)	934	UCU	1
dis3	3'-5' exoribonuclease subunit Dis3	1537	CUU	2
dld1	dihydrolipoamide dehydrogenase Dld1	456	UAU	3
ekc1	protein phosphatase regulatory subunit Ekc1 (predicted)	1972	UUG	1
elp1	elongator subunit Elp1 (predicted)	1859	GUU	2
elp2	elongator complex subunit Elp2 (predicted)	1101	GUA	2
eng1	endo-13-beta-glucanase Eng1	945	GAU	3
eng1	endo-13-beta-glucanase Eng1	275	UUA	2
erg27	3-keto sterol reductase Erg27 (predicted)	401	UUU	3
fhn1	plasma membrane organization protein Fhn1	66	GAU	3
fhn1	plasma membrane organization protein Fhn1	426	GUU	3
fhn1	plasma membrane organization protein Fhn1	1281	NA	NA
fim1	fimbrin	1786	CUU	3
fip1	iron permease Fip1	31	UUC	1
fra1	iron responsive transcriptional regulator peptidase family (predicted)	1429	ACU	3
gas2	13-beta-glucanosyltransferase Gas2 (predicted)	-31	NA	NA
glo3	ARF GTPase activating protein (predicted)	823	GCU	3
gls2	glucosidase II alpha subunit Gls2	1845	GUU	3
gls2	glucosidase II alpha subunit Gls2	675	GAU	3
gpd2	glycerol-3-phosphate dehydrogenase Gpd2	477	UCU	3
gpt1	UDP-glucose-glycoprotein glucosyltransferase Gpt1	2989	UUC	1
grs1	mitochondrial and cytoplasmic glycine-tRNA ligase Grs1	1597	UCU	1
hhp1	serine/threonine protein kinase Hhp1	446	AUA	2
his7	phosphoribosyl-AMP cyclohydrolase/phosphoribosyl- ATP pyrophosphohydrolase His7	24	GAU	3
hta1	histone H2A alpha	-223	NA	NA
hta1	histone H2A alpha	-258	NA	NA
kap104	karyopherin Kap104	776	GAU	3
lys2	homoaconitate hydratase Lys2	793	UUG	1

Supplemental Table 1 →

gene	description	ATG	codon	position
lys3	saccharopine dehydrogenase Lys3	2	AUG	2
mae2	malic enzyme malate dehydrogenase (oxaloacetate decarboxylating) Mae2	152	UUG	2
mms2	ubiquitin conjugating enzyme Mms2	877	NA	NA
msy1	MS calcium ion channel protein Msy1	2535	ACU	3
mug157	conserved protein Mug157	346	UCU	1
mug161	CwfJ family protein splicing factor (predicted)	1857	UUA	2
nda2	tubulin alpha 1	1133	GUU	2
nda2	tubulin alpha 1	1477	NA	NA
nda3	tubulin beta Nda3	803	UCU	1
nop14	U3 snoRNP protein Nop14 (predicted)	245	GUU	2
npl4	Cdc48-Npl4-Ufd1 complex subunit Npl4 (predicted)	740	UCU	3
nro1	negative regulator of Ofd1 Nro1	-1	NA	NA
nrs1	cytoplasmic asparagine-tRNA ligase Nrs1 (predicted)	1050	CAU	3
nuc1	DNA-directed RNA polymerase I complex large subunit Nuc1	389	CUU	2
pap1	transcription factor Pap1/Caf3	969	GUU	3
pdi1	protein disulfide isomerase (predicted)	777	ACU	3
pgr1	mitochondrial glutathione reductase Pgr1	1146	AUA	2
pik1	1-phosphatidylinositol 4-kinase Pik1	1797	AUU	2
plb1	phospholipase B homolog Plb1	36	UUU	3
ppp1	pescadillo-family BRCT domain protein Ppp1 (predicted)	1550	GUU	2
prp2	U2AF large subunit (U2AF-59)	41	AUA	2
ptc2	protein phosphatase 2C Ptc2	99	CGU	3
pub1	HECT-type ubiquitin-protein ligase E3 Pub1	1439	UAU	1
pup1	20S proteasome complex subunit beta 2 (predicted)	737	UGU	3
rar1	cytoplasmic methionine-tRNA ligase Mrs1 (predicted)	808	ACU	3
rhp23	Rad23 homolog Rhp23	233	CUU	2
rmn1	RNA-binding protein	552	UUG	1
rmn1	RNA-binding protein	349	GUU	2
rpl2502	60S ribosomal protein L25 (predicted)	-10	NA	NA
rrb1	WD repeat protein Rrb1 (predicted)	501	GAU	3
rrp5	U3 snoRNP-associated protein Rrp5 (predicted)	3855	GGU	3
sal3	karyopherin Sal3	2614	UAC	1
sam1	S-adenosylmethionine synthetase	1412	NA	NA
scl1	20S proteasome complex subunit alpha 1 (predicted)	-26	NA	NA
sds23	PP2A-type phosphatase inhibitor Sds23/Moc1	434	UUU	2
sec17	alpha SNAP (predicted)	452	AUU	2
sec21	coatomer gamma subunit Sec21 (predicted)	2009	UUU	3
ser2	phosphoserine phosphatase Ser2 (predicted)	890	CUU	3
sfp1	transcription factor Sfp1 (predicted)	204	GAU	3
slc1	1-acylglycerol-3-phosphate O-acyltransferase Slc1 (predicted)	8	UUU	2
smb1	Sm snRNP core protein Smb1	441	UAU	1
SPAC1635.01	mitochondrial outer membrane voltage-dependent anion-selective channel	519	CCU	3
SPAC1639.01c	GNS1/SUR4 family protein (predicted)	2140	NA	NA
SPAC1786.02	phospholipase (predicted)	372	CGU	3
SPAC17G6.03	phosphoprotein phosphatase (predicted) **position overlapping a tRNA**	-1089	NA	NA
SPAC17G6.03	phosphoprotein phosphatase (predicted) **position overlapping a tRNA**	-1092	NA	NA
SPAC1F5.03c	FAD-dependent oxidoreductase involved in late endosome to Golgi transport	1170	AAU	3
SPAC26H5.07c	seven transmembrane receptor protein (predicted)	711	UAC	1
SPAC56E4.03	aromatic aminotransferase (predicted)	1078	UAU	1
SPAC589.06c	pho88 family protein (predicted)	541	ACU	3
SPAC694.02	DEAD/DEAH box helicase	2303	UUU	2
SPAC694.02	DEAD/DEAH box helicase	2304	UUU	3
SPAC977.12	L-asparaginase (predicted)	1117	NA	NA
SPAPB2C8.01	cell surface glycoprotein adhesion molecule (predicted)	3009	GCU	3
spb1	rRNA methyltransferase Spb1 (predicted)	828	GUA	2

Supplemental Table 1 →

gene	description	ATG	codon	position
SPBC119.09c	ORMDL family protein (predicted)	773	NA	NA
SPBC119.17	mitochondrial metalloendopeptidase (predicted)	769	UUC	1
SPBC1271.10c	transmembrane transporter (predicted)	1687	UUA	1
SPBC12C2.11	glutamine-fructose-6-phosphate transaminase (predicted)	1502	AUC	2
SPBC1E8.05	conserved fungal protein	126	AAU	3
SPBC354.04	Schizosaccharomyces specific protein	-16	NA	NA
SPBC3F6.01c	serine/threonine protein phosphatase (predicted)	1206	CGU	3
SPBC3H7.03c	2-oxoglutarate dehydrogenase (lipoamide) (predicted)	1539	UCU	3
SPBC409.08	spermine family transmembrane transporter (predicted)	104	NA	NA
SPBC530.11c	transcription factor zf-fungal binuclear cluster type (predicted)	2233	GUG	2
SPBC646.08c	oxysterol binding protein (predicted)	1356	UAU	3
SPCC1739.01	zf-CCCH type zinc finger protein	377	UUA	2
SPCC4B3.03c	mitochondrial morphology protein (predicted)	806	CUU	2
SPCC550.11	karyopherin (predicted)	1310	UUU	1
SPCC553.10	conserved fungal protein	21	UUU	3
SPCC584.01c	sulfite reductase NADPH flavoprotein subunit (predicted)	223	UUG	1
SPCC584.11c	Svf1 family protein Svf1	1433	CUU	2
SPCC584.11c	Svf1 family protein Svf1	584	UUC	2
spp27	RNA polymerase I upstream activation factor complex subunit Spp27	711	UUG	2
tif471	translation initiation factor eIF4G	588	AGU	3
tif471	translation initiation factor eIF4G	3720	GAU	3
tim23	TIM23 translocase complex subunit Tim23 (predicted)	96	GAU	3
toa1	transcription factor TFIIA complex large subunit Toa1 (predicted)	489	UUU	1
trp3	anthranilate synthase component I (predicted)	483	GGU	3
trr1	thioredoxin reductase Trr1	881	AUU	3
trr1	thioredoxin reductase Trr1	560	GUU	3
ubx3	UBX domain protein Ubx3 Cdc48 cofactor	1127	UUC	2
usp107	U1 snRNP-associated protein Usp107	384	UUU	2
usp107	U1 snRNP-associated protein Usp107	1262	UUG	1
utp3	U3 snoRNP-associated protein Utp3 (predicted)	266	UUA	2
uvi15	UV-induced protein Uvi15	-54	NA	NA
vma2	V-type ATPase V1 subunit B	839	UUA	2
zwf1	glucose-6-phosphate 1-dehydrogenase (predicted)	1091	AUU	3

→ **Supplemental Table 1** – **List of mRNAs with at least one D-site in fission yeast.** 143 D-sites were detected on 125 products of protein-coding genes (gene column). The known or predicted function of the gene is depicted in the second column (description). The position of the D-site compared to the ATG start codon is indicated in the third column (ATG), where A is the nucleotide 1. A negative number means that the D-site was located on the 5'UTR, a positive number implies that the D-site was either on the coding sequence, or on an intron, or on the 3'UTR. When the D-site was found on a codon, the line is shaded, the corresponding codon is indicated (codon) and the position on the codon (5'-123-3') is precised (position). NA stands for *not applicable*.

gene	description
SPNCRNA.113	noncoding RNA

Supplemental Table 2 – **Noncoding RNA that is not a tRNA and that carries a D-site.** The lncRNA *SPNCRNA.113* overlaps with the 3'UTR of the *fhn1* protein-coding gene (see Table S1). The D-seq pipeline did not allow to discriminate which RNA (*fhn1* or *SPNCRNA.113*) was actually dihydrouridylated.

gene	description	codon	position
ags1	alpha glucan synthase Ags1	UGG	1
bms1	GTP binding protein Bms1 (predicted)	CUU	3
cct1	chaperonin-containing T-complex alpha subunit Cct1	CUU	2
dld1	dihydrolipoamide dehydrogenase Dld1	UAU	3
fim1	fimbrin	CUU	3
gas2	13-beta-glucanosyltransferase Gas2 (predicted)	NA	NA
hta1	histone H2A alpha	NA	NA
lys2	homoaconitate hydratase Lys2	UUG	1
nop14	U3 snoRNP protein Nop14 (predicted)	GUU	2
ptc2	protein phosphatase 2C Ptc2	CGU	3
rhp23	Rad23 homolog Rhp23	CUU	2
ser2	phosphoserine phosphatase Ser2 (predicted)	CUU	3
slc1	1-acylglycerol-3-phosphate O-acyltransferase Slc1 (predicted)	UUU	2
SPAC17G6.03	phosphoprotein phosphatase (predicted) **position overlapping a tRNA**	NA	NA
SPAC17G6.03	phosphoprotein phosphatase (predicted) **position overlapping a tRNA**	NA	NA
SPBC119.17	mitochondrial metalloendopeptidase (predicted)	UUC	1
SPBC3F6.01c	serine/threonine protein phosphatase (predicted)	CGU	3
SPCC4B3.03c	mitochondrial morphology protein (predicted)	CUU	2
tif471	translation initiation factor eIF4G	AGU	3
zwf1	glucose-6-phosphate 1-dehydrogenase (predicted)	AUU	3

Supplemental Table 3 – **List of protein-coding genes with a detected D-site in both D-seq analyses with or without including the single *dus* mutants (see Addendum).** 20 D-sites were detected on 19 different products of protein-coding genes in both analyses (gene column). The known or predicted function of the gene is depicted in the second column (description). When the D-site was found on a codon, the line is shaded, the corresponding codon is indicated (codon) and the position on the codon (5'-123-3') is precised (position). NA stands for *not applicable* for D-sites on 5' or 3'-UTRs.